Understanding the Film Audience – Providing Insight into the Viewer's Experience from Text Mining and Manual Text Analysis of Online Film Reviews

Urszula Świerczyńska-Kaczor*, **Jacek Wachowicz****

The paper aims to contribute to the research of the film market, starting a discussion and seeking answers to the following problem: What spectrum of film-viewer experiences can be identified and better understood due to the implementation of text mining in the analysis of online film reviews? The presented study was based on the analysis of online audience reviews of five films targeted at a young audience, with their premières in 2016 and 2017. The findings suggest that implementing text mining as a method of analysis of online reviews can provide valuable insight into the film market, which may be helpful for producers in developing future productions, or altering the communication strategy.

Keywords: text mining, viewer research, film market.

Zrozumieć widownię filmową – badanie doświadczeń widzów za pomocą *text mining* oraz manualnej analizy internetowych recenzji filmów

Przedstawione w artykule rozważania mają na celu wzbogacenie obszaru badań dotyczących filmu oraz stanowią próbę znalezienia odpowiedzi na pytanie, czy spektrum doświadczeń widza może być zidentyfiko-wane oraz lepiej rozumiane poprzez zastosowanie *text mining* w analizie internetowych recenzji filmowych. Niniejsze badanie zostało oparte na analizie internetowych recenzji pięciu filmów kierowanych do młodej widowni, których premiery odbyły się w 2016 oraz w 2017 roku. Prezentowane wyniki badania pozwa-lają na wysnucie wniosku, iż zastosowanie *text mining* w analizie recenzji online pozwala na uzyskanie informacji przydatnych w zrozumieniu rynku filmowego, a także użytecznych dla producentów filmowych w przygotowywaniu przyszłych produkcji, lub też w korekcie obranych strategii komunikacji marketingowej.

Słowa kluczowe: text mining, badania widowni, rynek kinowy.

JEL: M300

* **Urszula Świerczyńska-Kaczor** – dr hab. Lodz Film School.

** **Jacek Wachowicz** – dr, UTP University of Science and Technology.

Correspondence address: Lodz Film School, Targowa 61/63 St., 90-323 Łódź, Poland; e-mail: uswier-czynska@filmschool.lodz.pl; UTP University of Science and Technology, Fordońska 430 St., 85-790 Bydgoszcz, Poland; e-mail: jacek.wachowicz@utp.edu.pl.

Introduction

Internet film forums play a crucial role in shaping film-viewers' experience nowadays, as the generated word-of-mouth (WOM) about film productions delivers information in the pre-purchase stage, and allows the audience to share the experience with others after watching a film. This article is focused on the word-of-mouth spread by 'non-professional' film reviewers whose voice – due to the ubiquity of the Internet – has become more prominent compared to the pre-Internet era. It may be presumed that the analysis of online WOM (textual data) can offer the researchers and film industry a valuable insight into the viewer's experience connected with a film production.

In this article, we look for answers to the following research question: What spectrum of film-viewer experiences can be identified and better understood due to the implementation of text mining techniques in the analysis of online film reviews? The discussion and results of the empirical study presented in this article are focused on watching films at the cinema (as opposed to watching films online, or via a TV channel, which can trigger different experiences). The analysis refers to the films which targeted young audiences: 'Sing' (2016), 'Alice Through The Looking Glass' (2016), 'Beauty and the Beast' (2017), 'Baby Boss' (2017), and the Oscar-winning animation 'Zootopia' (2016).

The article is structured as follows. The next part of the article highlights selected aspects of the viewer experience. It is followed by the presentation of the research framework. In point 3, the results of the conducted analysis are presented. However, it is worth emphasizing that due to the volume of results of text mining analysis, the presentation of the full research report is not possible – the article embraces presentations of selected parts of the research report. The conclusions, limitations and future research are presented in the last part of article.

The presented study is a part of a broader research project, conducted at Państwowa Wyższa Szkoła Filmowa, Telewizyjna i Teatralna: 'Badania statutowe, number WOSF/UPB/2017/09'.

1. Literature Background

A. Viewer Experience Construct

Film-viewer experience is a highly complex construct, and is built through three different stages: the expectations connected with viewing the film at the cinema, the process of experiencing the product at the cinema during its screening time, and the process of evaluation after consuming such experiential goods as films.

The expectation confirmation theory indicates that consumers' satisfaction depends on the relations between consumers' expectations and their perception of an actually delivered product. The well-known Service Quality Model indicates that customers' expectations are based on (Parasuraman, Berry and Zeithaml, 1991):

1. personal needs. Taking into account the viewers' needs, Polish cinemagoers can be divided into two major segments: 'Going to the cinema' or 'Watching the film'. The first segment values 'having free time at the cinema', while the second segment is more 'film-value' oriented (Wolny, Jaciow, Krężołek, Gadomska and Wawro, 2015);

2. word-of-mouth communication;

3. past experiences. Past experiences can be connected to the expectations of the genre by the viewer (the style of directing, the type of plot and so on). As to this factor, on the Polish family and children's movie market, the most popular 'children's film' genres are animations, followed by 'live-action', and a small, but growing number of documentary films (Gawron-Jaksik and Materska-Samek, 2016).

The next stage of forming the viewer experience – consumption of a product – depends (at least) on the following aspects:

1. The purchase decision. The child or children can significantly influence the decision on going to the cinema and watching a particular film production;

2. Social factors. Contrary to watching a film on the Internet, or experiencing Virtual Reality, watching a film at the cinema is always a social experience, and the viewer is affected by the behavior, and profile of others;

3. The cinema facilities. In the process of choosing a cinema, the Polish segment of cinemagoers "Watching the film" valued, among other aspects, 'comfort of watching' (25.1%), 'sale promotions offered at the cinema' (20.4%), and 'convenient access to parking space' (12.4%) (Wolny, Jaciow, Krężołek, Gadomska and Wawro, 2015).

All of the above listed aspects can become the subject of consumer evaluation of the product and experience and can be shared with others at the post-purchase phase, building a positive or negative WOM.

B. Word-of-Mouth: The Significance for the Movie Market

The research questions referring to characteristics of WOM, the factors influencing spreading and the persuasiveness of WOM have been the topic of many studies (for example, Chen and Xie, 2008; Anderson, 1998; Hu, Pavlou and Zhang, 2006; West and Broniarczyk, 1998; Robinson, Goh and Zhang, 2012; Hsieh, Hsieh and Tang, 2012; Bickart and Schindler, 2001; Bambauer-Sachse and Mangold, 2011; Herr, Kardes and Kim, 1991). As to the film market, online reviews have become vital for the movie industry. Eliashberg and Shugan (1997) found that critical reviews of professional film

critics were uncorrelated with early box office results, but critical reviews tend to correlate with the total box office revenue. The findings of Basuroy, Chatterjee and Ravid (2003) indicate that both positive and negative reviews correlated with weekly box office revenue; the effect of negative reviews diminished over time; the effect of negative reviews was more significant than positive reviews over the first week; and that box office revenue was moderated by popular stars and big budgets. The research of Belvaux and Marteaux (2007) shows that promotion, professional film critics and web-users reviews affected movie attendance.

C. Understanding Viewer Experience

'Traditional' methods of understanding the film-viewer's expectations and experiences (for example individual interviews or focus groups) are often expensive, time-consuming or geographically limited. An alternative method is to look for relevant data on movie-related portals, such as the international and well-known website Rotten Tomatoes®, or the Polish Filmweb. In this situation, the process of analysis is based on data in a form of user-generated online reviews being written as informal text, often with spelling and grammar mistakes and very different styles, depicting the viewer's experience with a film.

Analyses of text data available online as film reviews can be conducted manually. However, the manual online review assessment has its disadvantages, such as the data volume to process (large data volume leads to the necessity of sampling), complexity of data coding and time limitations. A possible alternative approach lies in the implementation of text mining techniques. The approach based on text mining allows the extraction of meaningful patterns from large volumes of textual information (Turban, Sharda and Delen, 2010), leading to receiving a 'quick overview' of the investigated subject, although the more detailed analysis often needs manual work (Kayser and Blind, 2017). A deeper information processing technique is opinion mining, which captures the subjectivity in terms of the semantic orientation associated with the constituents of a text (Gräbner, Zanker, Fliedl and Fuchs, 2012; Taboada, Brooke, Tofiloski, Voll and Stede, 2011).

2. The Research Framework

In the study conducted in March–April 2017, we investigated the following research questions: What spectrum of film-viewer experiences can be identified and better understood due to the implementation of text mining techniques in the analysis of online film reviews?

The study was structured as follows:
1. Gathering data from online reviews. Five films, produced by major film studios and screened in 2016 and 2107, were selected for analysis: 'Zootopia' (2016), 'Sing' (2016), 'Alice Through The Looking Glass'

(2016), 'Beauty and the Beast' (2017), and 'Baby Boss' (2017) (Table 1). All selected films are rated as PG – Parental Guidance Suggested. We analyzed the posts of film audiences (not professional film critics) submitted by Internet users to the popular film website Rotten Tomatoes® – rottentomatos.com. In April 2017 a random sample of posts written in English (other languages excluded) was obtained: 'Zootopia' – 702 posts, 'Sing' – 599 posts, 'Alice Through the Looking Glass' – 935 posts, 'Beauty and the Beast' – 917 posts, 'Baby Boss' – 365 posts (sample description – Table 2).

2. Data analysis. To analyze the data, we used the text mining software Statistica and KH Coder. We also looked for an explanation and relations between extracted words and their meaning, conducting a manual analysis of sub-samples (see Table 2). As it is not possible to present the whole spectrum of results extracted from text mining due to their volume or graphical forms and the limited space of this article, in the presentation of findings we put not all, but selected graphs and tables for the analyzed film productions.

Nr	Title	Information	Studio	Rating for the audience Genre
1.	Zootopia (2016) Première: world: 10.02.2016 Poland: 19.02.2016	Directed by: Byron Howard, Rich Moore, Jared Bush	Walt Disney Animation Studios	PG (Parental Guidance Suggested) – Action & Adventure, Animation, Comedy
2.	Sing (2016) Première: world: 11.09.2016 Poland: 06.01.2017	Directed by: Garth Jennings, Christophe Lourdelet	Illumination Entertainment	PG – Animation, Comedy, Musical & Performing Arts
3.	Alice Through The Looking Glass (2016) Première: world: 10.05.2016 Poland: 26.05.2016	Directed by: James Bobin	Walt Disney Pictures	PG – Action & Adventure, Kids & Family, Science Fiction & Fantasy
4.	Beauty and the Beast (2017) Première: world: 23.02.2017 Poland: 17.03.2017	Directed by: Bill Condon	Walt Disney Pictures	PG – Drama, Musical & Performing Arts, Science Fiction & Fantasy, Romans
5.	Baby Boss (2017) Première: world: 10.03.2017 Poland: 21.04.2017	Directed by: Tom McGrath	DreamWorks Animation	PG – Animation, Comedy, Kids & Family

Tab. 1. Description of the films selected for analysis. Source: http://www.filmweb.pl [02.04.2017], https://www.rottentomatoes.com/ [02.04.2017].

Urszula Świerczyńska-Kaczor, Jacek Wachowicz

No.	Title	Analysis based on software					Manual text analysis	
		Sample	Average rating from obtained sample (0.5–5.0)	Standard deviation	Median	Mode	Sample	Analyzed posts indicating directly watching film 'with child/children'
1.	Zootopia (2016)	N1 = 702	4.28 (n = 698)	1.09	4.50	5.00	87 posts	3
2.	Sing (2016)	N2 = 599	Not calculated					7
3.	Alice Through The Looking Glass (2016)	N3 = 935	3.26 (n = 935)	1.39	3.50	5.00	31 posts	9
4.	Beauty and the Beast (2017)	N4 = 917	3.86 (n = 901)	1.27	4.00	5.00	91 posts	2
5.	Baby Boss (2017)	N5 = 365	3.33 (n = 365)	1.61	4.00	5.00	57 posts	11

Tab. 2. Descriptive statistics – sample of audience reviews selected from Rotten Tomatoes®. Source: authors' own analysis, the posts selected from https://www.rottentomatoes.com, date of sample selection 16–23.04 2017, note – the date of writing posts by viewer can be different than the date of sample selection.

3. Results

3.1. Singular Value Decomposition (SVD) of Analyzed Text: Extraction of Concepts, Word Importance, Word Residual, SVD Word Coefficients

Singular value decomposition led to the extraction of the concepts – which are 'a few underlying dimensions that account for most of the common contents or meaning of the documents and words that were extracted' (Statistica documentation – http://documentation.statsoft.com, access: 20.03.2017). Table 3 presents the number of concepts and top concepts extracted for the five analyzed films. For all analyzed films, two or three concepts explained most of the singular values (see – Figure 1). Table 4 and Table 5 present the top twenty words of the word importance values and the word residuals. Apart from the general word 'movie/film' and specific words connected directly with the plot (for example: 'babi', 'beast', 'tim', 'judi'), the analysis captured words depicting the viewers' sentiment towards a particular production:

1. Words with positive meaning: love, fun/funny, great, good, amazing, cute, enjoy, best, awesome, hilarious;
2. Words with negative meaning, such as 'bad';

The next step of the analysis was based on examining the scatter plots of word coefficients: concept 1 against concept 2 or concept 3. This analysis helped to understand the meaning of the semantic structure of the analyzed texts due to graphically 'capturing' the meaningful clusters of words (see – Figure 2).
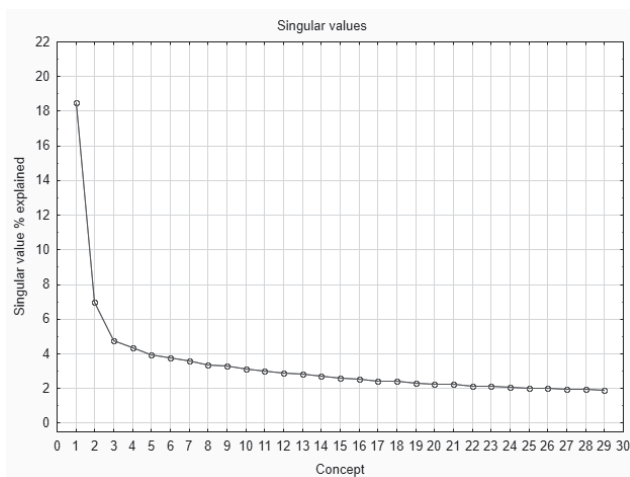

Fig. 1. Zootopia (2016) – Screen plot for singular values (Statistica soft ware)

Zootopia Number of concepts = 29		Sing Number of concepts = 33		Alice Through the Looking Glass Number of concepts = 28		Beauty and the Beast (2017) Number of concepts = 26		Baby Boss (2017) Number of concepts = 34	
	Value		Value		Value				
Concept 1	111.70	Concept 1	62.33	Concept 1	141.11	Concept 1	113.19	Concept 1	75.18
Concept 2	42.26	Concept 2	31.41	Concept 2	54.12	Concept 2	41.70	Concept 2	33.30
Concept 3	28.93	Concept 3	24.56	Concept 3	34.84	Concept 3	30.89	Concept 3	21.43
Concept 4	26.20	Concept 4	20.38	Concept 4	31.58	Concept 4	28.85	Concept 4	20.13
Concept 5	23.68	Concept 5	18.49	Concept 5	29.27	Concept 5	27.18	Concept 5	19.53

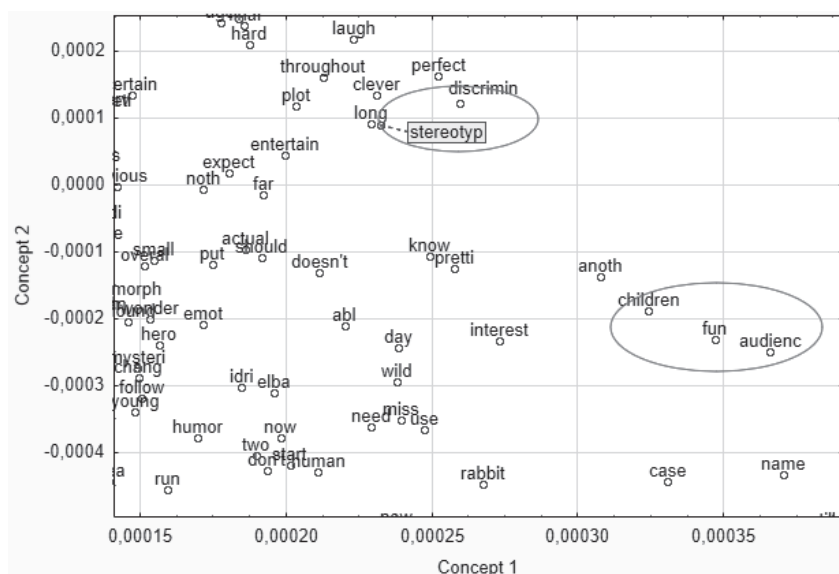Tab. 3. Singular values for analyzed films (Statistica software) – top five concepts

Fig. 2. Zootopia (2016) – Excerpt of the scatter plot of SVD word coefficients (Statistica software)

	Zootopia	Value	Sing	Value	Alice Through the Looking Glass	Value	Beauty and the Beast (2017)	Value	Baby Boss (2017)	Value
1	movi	100.00	movi	100.00	movi	100.00	movi	100.00	babi	100.00
2	anim	50.48	sing	72.92	alic	64.12	beast	64.99	movi	83.14
3	zootopia	45.46	film	61.94	time	52.59	bell	47.89	boss	46.71
4	judi	33.25	anim	60.05	stori	34.34	origin	46.56	tim	34.24
5	disney	29.53	charact	49.28	charact	33.83	like	41.00	anim	33.34
6	great	25.10	voic	48.59	one	31.43	beauti	40.33	voic	29.72
7	one	23.20	good	45.58	like	31.38	disney	36.77	kid	29.37
8	stori	22.58	great	44.31	look	28.35	love	36.68	like	28.89
9	charact	21.90	kid	37.32	just	25.86	good	35.25	realli	27.69
10	like	21.32	music	34.82	first	25.54	anim	33.28	funni	23.91

Tab. 4. Top ten words according to the word importance (Statistica software)

	Zootopia	Value	Sing	Value	Alice Through the Looking Glass	Value	Beauty and the Beast (2017)	Value	Baby Boss (2017)	Value
1	fun	66.52	cute	32.53	watch	108.97	see	96.74	thought	20.92
2	realli	55.31	expect	31.17	enjoy	98.01	better	96.47	fun	18.32
3	love	53.00	better	30.55	much	92.23	scene	95.56	hilari	16.62
4	funni	52.70	famili	28.17	see	85.36	amaz	92.95	bad	15.86
5	well	49.28	amaz	26.84	fun	85.27	best	92.07	cute	15.78
6	famili	49.25	will	26.65	great	82.79	watch	88.60	watch	15.09
7	see	46.75	lot	26.58	better	79.65	enjoy	86.11	lot	15.06
8	like	45.91	much	25.98	effect	78.00	much	84.64	children	14.46
9	just	44.79	laugh	25.82	origin	74.81	look	82.60	best	14.26
10	much	43.61	adult	25.77	plot	71.31	cast	82.02	adult	14.05

Tab. 5. Top ten words according to the word residual (Statistica software)

3.2. Hierarchical Cluster Analysis of Words, Multi-Dimensional Scaling of Words, Correspondence Analysis of Words – Analysis Based on KH Coder

Further analysis was conducted by implementing KH Coder, including: the hierarchical cluster analysis of words, multi-dimensional scaling of words, and the correspondence analysis of words (see Table 6, Figure 3, Figure 4).

Title	Selected clusters from hierarchical cluster analysis (note – the presentation of full results of this analysis in graphical forms is not possible)
Zootopia (2016)	'Disney-best', 'really-good', 'message-movie-great', 'lesson-teach', 'racism-prejudice', 'kid-adult-enjoy', 'family-time-watch', 'entertaining-funny-way', 'today-relevant', 'laugh-theme', 'young-audience', 'cute-funny-perfect', 'recommend-truly-deserve', 'definitely – favorite'
Sing (2016)	'fun-family-love', 'really-enjoy', 'funny-cute', 'great-movie-good', 'best-theater-animated', 'laugh – pretty', 'definitely-plot-predictable' 'enjoyable-entertaining'
Alice Through the Looking Glass (2016)	'enjoy-think-really-like', 'great-love-movie-good', 'plot-little', 'sequel – original', 'performance-best', 'family-fun', 'director-James Bobin', 'fantastic-creative-impossible', 'scene – worst', 'point-script', 'impressive – role – hope', 'enjoyable – overall', 'visually – stunning', 'storyline – boring – terrible', 'definitely – recommend', 'beautiful – absolutely', 'amazing – entertaining'
Beauty and the Beast (2017)	'character-feel', 'really-enjoy', 'love-movie-good', 'new-song', 'like-lot-favorite', 'better-role', 'moment-gay', ''live-action'-classic', 'best-year', 'better-expect', 'bad-especially', 'magical-time-tale-old', 'job-great', 'kid-amazing', 'child-recommend', 'fantastic-performance-singing', 'star-wonderful-absolutely', 'addition-perfect', 'plot-change'
Baby Boss (2017)	'family-enjoy', 'love-fun-hilarious', 'definitely-recommend', 'humor-kid-adult', 'funny-movie', 'really-good', 'DreamWorks-film', 'expect-way-better', 'grate-animation-story', 'laugh-theater', 'enjoyable-year-old', 'bad-end'

Tab. 6. Hierarchical cluster analysis – presentation of selected clusters extracted from analysis (KH Coder software)
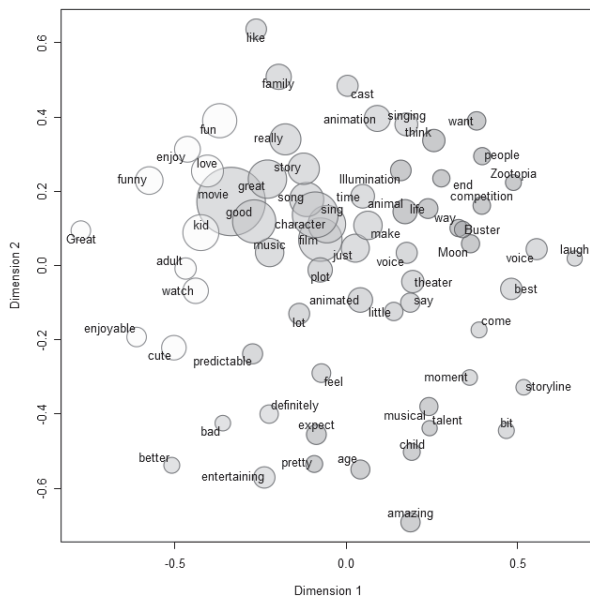
Fig. 3. Sing (2017) – multi-dimensional scaling of words, (KH Coder software)
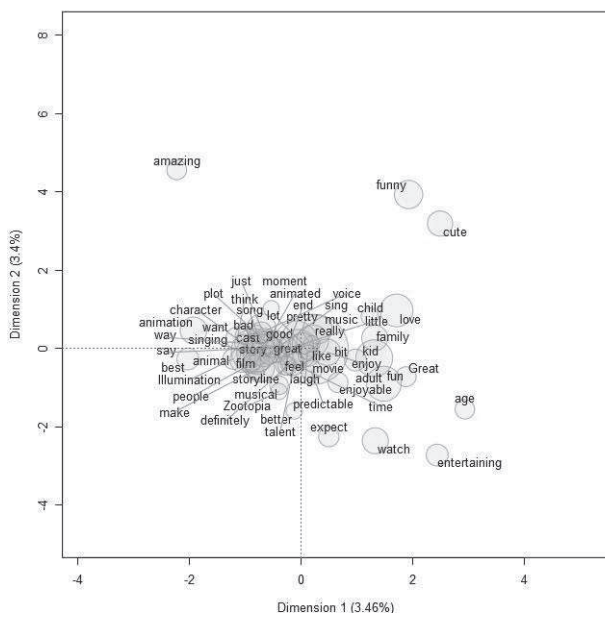

Fig. 4. Sing (2017) – correspondence analysis of words (KH Coder software)

3.3. Which Words of Online Reviews Are Predictors of Film Ranking?

Within the obtained samples of reviews, the highest audience rating received was 'Zootopia', with an average viewer-liking: 4.28 (this high rating is not surprising as this film was an Oscar winner in 2017 in the category 'Animated Feature Film'), followed by 'Beauty and the Beast' – 3.86, 'Baby Boss' – 3.33 and 'Alice through the Looking Glass' – 3.26 ('Sing' – was not processed due to the rating).

By implementing the Statistica text mining software, we identified words which were predictors of the rating given by the viewer (Table 7). The results show that words, for example – 'waste', 'aw/awful', 'hilarious', 'horrible', overrated', 'bore', 'hate', 'dull', 'ruin' – are among the best predictors of the film rating ($p < 0.02$).

The posts on 'Baby Boss' were selected from the Rotten Tomatoes® (rottentomatos.com) website before the film had its première in Poland (on 21st April 2017). Therefore, the next interesting aspect of analysis is whether online reviews (posts written before the Polish film première) were predictors of the first film box office revenue in Poland. Looking into relations between the number of viewers in Polish cinemas for 'Boss Baby' during the first days of screening (including the first weekend) and our text mining analysis, we cannot point to the specific factors which explain the ticket sales (Table 8).

Zootopia (2016)		Alice Through the Looking Glass (2016)		Beauty and the Beast (2016)		Baby Boss (2017)	
Predictor	Value F	Predictor	Value F	Predictor	Value F	Predictior	Value F
overr	14.95	aw	28.02	horribl	53.63	half	21.05
bore	12.07	bare	16.38	bore	24.21	wast	19.29
pay	10.22	materi	16.36	money	23.84	hilari	15.97
hate	9.02	care	15.68	terribl	16.57	bore	15.24
came	6.93	love	15.57	forc	15.88	worst	10.19
base	6.42	bore	15.06	ruin	13.39	hard	10.15
prove	6.24	ask	14.72	leav	13.36	horribl	9.66
best	5.49	suck	13.86	poor	12.64	hour	9.66
color	5.43	script	13.62	fell	10.67	i'll	8.68
can't	5.43	man	13.07	togeth	9.90	bad	8.63
tri	5.20	live	13.00	absolut	9.74	save	8.25
decent	5.11	dull	12.33	bad	8.68	leav	7.43
wasn't	5.10	money	11.92	miss	8.34	star	7.36
engag	4.62	wast	11.39	noth	8.12	terribl	7.33
goe	4.61	sourc	11.36	lack	7.31	money	6.42

Tab. 7. The best 15 predictors for rating (p<0.02, Statistica software)

	Date of première – Polish market	Date of first weekend of screening	Box office – first weekend	Number of viewers from the première – during 3 days	Number of film copies	Number of viewers per film copy
Baby Boss (2017)	21.04.2017	21–23.04.2017	207 021	281 910	294	959
Zootopia (2016)	19.02.2016	19–21.02.2016	230 833	230 833	233	991
Sing (2016)	06.01.2017	06–08.2017	255 410	287 161	250	1149
Alice Through the Looking Glass (2016)	26.05.2016	27–29.05.2016	58 117	77 053	230	335
Beauty and the Beast (2017)	17.03.2017	17–19.03.2016	230 800	230 800	264	874

Tab. 8. The opening box office of films on Polish market for analyzed films. Source: analysis based on the data presented on Stowarzyszenie Filmowców Polskich website, https:// www.sfp.org.pl/box_office [30.04.2017]

3.4. Manual Analysis and Further Interpretation of Text Mining Analysis

The text mining analysis and the manual text analysis led to the identification of factors influencing the viewer's experience. Therefore (see Table 9):

- Viewers' experiences varied significantly among the audience: the perception of the way in which the film was directed, the evaluation of quality of the cast, music, and special effects.
- The viewer's experience is framed by the expectation based on the perception of the studio brand.
- The viewer's experience is framed by the expectation based on previous films, the films which are screened at the same time in the cinema, and the general 'genre' of the film such as animation or 'live-action'.
- The viewer's experience is framed by political and social issues. Viewers find the 'second layer' within the film ('the layer beyond the adventure story').
- Viewer's experience is strongly framed by going to the cinema with the family. Viewers emphasize that they spend 'valuable time with family/daughter/son/grandson'. Reviews indicate that the parent/parents/ grandparent went to the cinema with children of a wide range of ages (from 3 years old to teenagers and grown-ups). The reviewer attitude is sometimes different from the child's/children's – 'I don't like the film, but my daughter / son / grandson / granddaughter likes the film'; posts within the sub-sample selected for manual analysis may suggest more positive attitudes in the reviews directly indicating watching the film

with a child or children compared to other reviews (this finding is not conclusive due to the very limited sample – see Table 2).

- A visit to the cinema is linked to the 'nostalgia' interpreted from two perspectives. 1. Parents remembered watching the film previously with small children, nowadays the same people watch the film together as grown-ups. 2. Adults remember watching the film as a child.
- Viewers would like to 'keep' the positive experience of watching the film at the cinema by buying a digital film copy.

Manual text analysis	Text mining results
'it is old Disney/DreamWorks/ Illumination', *'it was not what I expected from…'.*	SVD analysis – word importance shows names of studio brands; Hierarchical cluster analysis shows connections between brands, for example 'Disney-best'
Viewers directly compared the films with previous films: 'Alice in the Wonderland' (2010), 'Beauty and the Beast' (1991), *'the classic was better/ worse…'.*	SVD analysis – word importance and word residual – shows the words 'classic' and 'origin' among the most important for 'Beauty and the Beast' and 'Alice Through the Looking Glass'
Viewers of 'Zootopia' perceived the film as raising racism issues or political issues. Viewers of 'Beauty and the Beast' perceived the film as raising the homosexuality theme.	'Zootopia' – SVD: word importance – shows that among important words there is the word 'politic'; 'Zootopia' – Fig. 2. shows the cluster of words: discrimination, stereotype; 'Zootopia' – hierarchical cluster analysis shows connection: 'racism-prejudice'; 'Zootopia' – multi-dimensional scaling of words with adverbs and adjectives shows the 'political' element; 'Beauty and the Beast' – hierarchical cluster analysis – shows the link 'moment-gay'

Tab. 9. The selected results for manual text analysis and corresponding text mining results

Conclusions and Future Research

The research referring to the films for a young audience seems to be very limited. The study presented in this article aims to partly fill this gap, seeking the answers to the following research question: What spectrum of film-viewer experiences can be identified and better understood due to the implementation of the text mining techniques in the analysis of online film reviews?

In this article, we presented the analysis for the five selected films 'Zootopia', 'Sing', 'Alice Through the Looking Glass', 'The Beauty and the Beast', and 'Baby Boss'. Text mining analysis of online film reviews (written by the audience, not professional film critics) provided insight about the structure of textual data. Some words turned out to build a catalog of 'the important descriptive words for the film viewer's experience', and

for example, words – amazing, best, great, fun – frequently appeared as important in the text analysis. Therefore, the analysis of viewer experiences should be based both on:

1. Indicating which words appear among the results for a particular film title;
2. Indicating which 'obvious' words are missing, meaning – they do not appear in the analysis conducted for a particular title.

We labeled some words extracted from text mining as the 'peculiarities' – the words without an obvious connection with the plot, but they were significant for understanding the viewers' reviews. Such words can be indicators of the viewers' 'not obvious' frame for experiencing a particular film, for example through the lens of a political or social issue.

In the conducted study, some identified factors shaping viewer experiences can be easily predicted, such as a film studio brand, previous films, competitive productions, online reviews read on film websites by the viewer, family values connected with spending time together. Other results are surprising, such as strongly experiencing the film for a young audience through the lenses of political or social issues, and the feeling of 'nostalgia' – understood as 'feeling like a child again' or 'as a parent experiencing their kid's childhood'. Moreover, the preliminary findings suggest that reviews directly indicating 'going to the cinema with child/children' may be more positive overall than other reviews. It is also important for the viewer to 'keep' the positive experience of watching the film at the cinema by buying a digital film copy.

The conducted analysis shows that the best qualitative predictors of the viewer's rating have a negative meaning, for example: 'waste money', 'bore', 'worst film', 'awful', 'can't bear', 'horrible', 'miss (sense of plot)', 'leave (cinema/in the middle of screening)'. Negative words dominated the pool of the best predictors for film rating, and they are not linked to the general film rating. As to the predictive power of these words for opening box office results, the first insight presented in this article may suggest that there is no relation between the extracted words and first opening box office revenue for a new geographical market.

The study presented in this article has its limitations, such as:

1. Sample of films selected for analysis includes titles which aim at a young audience (as they are types of 'fairy-tales'), but on the other hand they are not homogenous in genre (animation, live-action);
2. The sample of selected films was small compared to the number of titles which can be categorized as a 'film for young audiences';
3. There is a possibility of the distortion of results due to the process of post sampling.

The presented study has managerial implications for the film industry. The study shows that the implementation of text mining software in the

process of the analysis of online reviews is a method which allows the film industry to quickly and cheaply indicate areas of the viewer experience which may need to be highlighted in their communication strategy, for example using social media. Secondly, using text mining methods provides insight about the film market, which may be helpful for producers in product development.

Future studies should give a more clear understanding of the relations between the results of text mining analysis and box office revenues. The interesting directions of future research are also:

1. Implementing text mining analysis as a tool for the evaluation of product placement. Many films depend on product placement as a source of financing costly productions. So far, the effectiveness of product placement is one of the most under-examined areas. Is it possible to use the text mining results of online reviews as indicators of the effectiveness of communication with prospective customers through product placement?

2. Building the model aiming to predict the early box office revenue and the cumulative revenue (e.g. form Video on Demand, DVD). Customers, on the basis of online film reviews, not only decide to watch the film at the cinema, but also they decide about buying film DVDs. Does the model including the results of text mining film reviews among other factors give a better evaluation of future cumulative film revenues compared to the models presented in literature so far?

References

Anderson, E.W. (1998). Customer satisfaction and word of mouth. *Journal of Service Research*, *1*(1), 5–17, https://doi.org/10.1177/109467059800100102.

Bambauer-Sachse, S. and Mangold, S. (2011). Brand equity dilution through negative online word-of-mouth communication. *Journal of Retailing and Consumer Services*, *18*(1), 38–45, https://doi.org/10.1016/j.jretconser.2010.09.003.

Basuroy, S., Chatterjee, S. and Ravid, S.A. (2003). How critical are critical reviews? the box office effects of film critics, star power, and budgets. *Journal of Marketing*, *67* (October), 103–117, https://doi.org/10.1509/jmkg.67.4.103.18692.

Belvaux, B. and Marteaux, S. (2007). Web user opinions as an information source. What impact on cinema attendances? *Recherche et Applications en Marketing* (English Edition) (AFM c/o ESCP-EAP), *22*(3), 65–81, https://doi.org/10.1177/205157070702200305.

Bickart, B. and Schindler, R.M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, *15*(3), 31–40, doi:10.1002/dir.1014.

Chen, Y. and Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, *54*(3), 477–491, http://dx.doi.org/10.1287/mnsc.1070.0810.

Eliashberg, J. and Shugan, S.M. (1997). Film critics: Influencers or predictors? *Journal of Marketing*, *61* (April), 68–78, http://dx.doi.org/10.2307/1251831.

Gawron-Jaksik, A. and Materska-Samek, M. (2016). *Czy kinematografia kręci się wokół dzieci? Kondycja kinematografii i analiza uwarunkowań edukacji filmowej dzieci do 12 roku życia*. Kraków: Fundacja Rozwoju Kina.

Gräbner, D., Zanker, M., Fliedl, G. and Fuchs, M. (2012). Classification of customer reviews based on sentiment analysis. In: *19th Conference on Information and Communication Technologies in Tourism* (pp. 460–470). Vienna, Austria: Springer, DOI: 10.1007/978-3-7091-1142-0_40.

Herr, P.M., Kardes, F.R. and Kim, J. (1991). Effects of word-of-mouth and product attribute information on persuasion: An accessibility-diagnosticity perspective. *Journal of Consumer Research*, *17*(4), 454–462, https://doi.org/10.1086/208570.

Hsieh, J.K., Hsieh, Y.C. and Tang, Y.C. (2012). Exploring the disseminating behaviors of eWOM marketing: persuasion in online video. *Electronic Commerce Research*, *12*(2), 201–224, http://dx.doi.org/10.1007/s10660-012-9091-y.

Hu, N., Pavlou, P.A. and Zhang, J. (2006). Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication. In: *Proceedings of the 7th ACM conference on Electronic commerce* (pp. 324–330).

Kayser, V. and Blind, K. (2017). Extending the knowledge base of foresight: The contribution of text mining. *Technological Forecasting & Social Change*, *116*, 208–215, http://dx.doi.org/10.1016/j.techfore.2016.10.017.

Parasuraman, A., Berry, L.L. and Zeithaml, V.A. (1991). Perceived service quality as a customer-based performance measure: An empirical examination of organizational barriers using an extended service quality model. *Human Resource Management* (Fall), *30*(3), 335–364, http://dx.doi.org/10.1002/hrm.3930300304.

Robinson, R., Goh, T. and Zhang, R. (2012). Textual factors in online product reviews: a foundation for a more influential approach to opinion mining. *Electronic Commerce Research*, *12*(3), 301–330, http://dx.doi.org/10.1007/s10660-012-9095-7.

Statistica documentation. Retrieved from: http://documentation.statsoft.com (20.03.2017).

Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, *37*(2), 267–307, http://dx.doi.org/10.1162/COLI_a_00049.

Turban, E., Sharda, R. and Delen, D. (2010). *Decision support and business intelligence systems* (9th ed.). Upper Saddle River, NJ: Prentice Hall.

West, P.M. and Broniarczyk, S.M. (1998). Integrating multiple opinions: The role of aspiration level on consumer response to critic consensus. *Journal of Consumer Research*, *25*(1), 38–51, http://dx.doi.org/10.1086/209525.

Wolny, R., Jaciow, M., Krężołek, D., Gadomska, M. and Wawro, A. (2015). *Widz kinowy w Polsce. Raport z badań.* Katowice: Uniwersytet Ekonomiczny w Katowicach, Centrum Badań i Transferu Wiedzy (August).