DOI http://dx.doi.org/10.1590/1981-5344/60611

ANALISANDO FILMES POR MEIO DE TAGS: A REPRESENTAÇÃO TEMÁTICA DE FILMES DE TERROR DA DÉCADA DE 1980 EM PLOT KEYWORDS E SINOPSES DO IMDB

Rafael Rodrigo do Carmo Batista
Universidade Federal de Santa Catarina (UFSC)
Florianópolis, Santa Catarina, Brasil

Thiago Henrique Bragato Barros
Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre, Rio Grande do Sul, Brasil

RESUMO

Introdução: a representação temática de filmes constitui um desafio para a Ciência da Informação, especialmente em bases colaborativas, nas quais usuários atribuem palavras-chave às obras. Objetivo: verificar em que medida as *plot keywords* atribuídas pelos usuários da *Internet Movie Database* (IMDb) representam o conteúdo temático de filmes de terror lançados entre 1980 e 1989, em comparação com os termos mais frequentes das sinopses. **Metodologia:** foram coletadas as 100 *plot keywords* mais votadas dos filmes mais populares do período e as 100 palavras mais frequentes nos resumos de enredo de 127 títulos recuperados por busca. Os termos foram distribuídos em seis dimensões de análise derivadas das definições do gênero terror. **Resultados:** as *plot keywords* contemplam todas as dimensões e nomeiam elementos específicos, enquanto as sinopses se concentram em ambientação e comunidade, com termos predominantemente genéricos. Parte das *keywords* refere-se a aspectos de produção ou a curiosidades. **Considerações finais**: as *plot keywords* oferecem uma representação mais específica, porém parcial, do conteúdo fílmico, e os dois conjuntos constituem formas complementares de representação da informação audiovisual.

Palavras-chave: Categorização de filmes. Representação da informação. Mineração de dados. Documentação Audiovisual.

ANALYZING FILMS THROUGH TAGS: THE THEMATIC REPRESENTATION OF 1980S HORROR FILMS IN IMDB PLOT KEYWORDS AND PLOT SUMMARIES

ABSTRACT

Introduction: the thematic representation of films is a challenge for Information Science, especially in collaborative databases, where users assign keywords to works. **Objective:** to examine to what extent the plot keywords assigned by users of the Internet Movie Database (IMDb) represent the thematic content of horror films released between 1980 and 1989, compared with the most frequent terms in their plot summaries. **Method:** the 100 most-voted plot keywords from the most popular films of the period and the 100 most frequent words in the plot summaries of 127 titles retrieved by search, mined with Cinemagoer and spaCy, were collected. The terms were distributed across six analytical dimensions derived from definitions of the horror genre. **Results**: plot keywords cover all dimensions and name specific elements, whereas the plot summaries concentrate on setting and community, with predominantly generic terms. However, some keywords refer to production aspects or trivia. **Conclusion:** plot keywords provide a more specific, yet partial, representation of film content, and the two sets are complementary forms of audiovisual information representation.

Keywords: Film categorization. Information representation. Data mining. Audiovisual documentation.

DOI http://dx.doi.org/10.1590/1981-5344/60611

DECLARAÇÃO DE DISPONIBILIDADE DE DADOS
data-available – Os dados de pesquisa estão disponíveis em repositório.
https://github.com/rafaelrbcarmo/ArtigoIMDB

Recebido em: 27/07/2025
Aceito em: 20/10/2026
Editora Chefe: Lorena Tavares de Paula

1 INTRODUÇÃO

O estudo de aspectos relacionados aos gêneros documentais audiovisuais vem recebendo cada vez mais a atenção dos cientistas da informação (Cordeiro, 1996). Suas características particulares impõem uma série de questões relacionadas à forma de se tratar e representar tematicamente o seu conteúdo, visando facilitar a posterior recuperação, principalmente em catálogos *online*. Soma-se a isso uma tendência, principalmente em uma atualidade dominada por serviços de *Video on Demand* (VOD), de que um tratamento mais eficiente de representação do documento audiovisual seja cada vez mais necessário, principalmente ao se considerar que o espectador utiliza estratégias distintas de busca desses artefatos em relação a outros gêneros documentais.

Tais mudanças são particularmente observadas na dinâmica entre telespectador e serviço como o VOD. As particularidades da pesquisa dentro de plataformas de *streaming* como *Netflix* [2] ou *Prime Video* [3] estão mais relacionadas ao tipo de informação do que à tradicional pesquisa por elementos técnicos (nome de diretor, elenco, data de lançamento e outros) de um filme. Os elementos técnicos ainda são importantes, pois também oferecem por si só uma ferramenta de individualização de filmes em um catálogo muitas vezes vasto, mas a influência da pesquisa em outras palavras-chave ou *tags* está diretamente associada, por exemplo, à forma com que a *Netflix* categoriza a pesquisa dentro da sua interface.

A tarefa de categorizar um filme dentro de um gênero ou associá-lo a uma determinada *tag* de sentimento se constitui, dado o exposto, como uma tarefa evidentemente complexa. Agregar novos recursos ao ato de categorizar filmes torna esta categorização mais rica e relevante, sendo o agrupamento temático uma meta a ser alcançada.

Para tratar de representação temática de filmes – um tipo específico do documento audiovisual – é preciso considerar os agrupamentos temáticos que já são destacados seja pelos espectadores quanto pela pesquisa em cinema – os gêneros cinematográficos. Em termos gerais, o gênero agrega características de enredo comuns a um conjunto de filmes em uma categoria. Por meio de

[2] *Site*: https://www.netflix.com/br/.
[3] *Site*: https://www.primevideo.com.

DOI http://dx.doi.org/10.1590/1981-5344/60611

categorias como terror, romance ou drama, o espectador consegue, em um primeiro contato, reconhecer em algum grau quais elementos estarão presentes antes de ter contato com o filme a ser visto.

A divisão de filmes em gêneros, contudo, não é definitiva. Dada a natureza difusa dos conteúdos presentes em um filme, a categorização básica em gêneros populares nem sempre compreende todas as possibilidades narrativas de um filme. Por isso, categorizar filmes é essencialmente uma tarefa complexa.

Com o surgimento e desenvolvimento das bases de dados na internet, logo surgiriam alternativas de aplicação para o cinema. Os dados sobre filmes incluem tanto informações técnicas, relevantes para uma indústria que movimenta valores consideráveis, quanto informações relacionadas à narrativa (*plot*) e sinopses. Em um sentido de representação da informação audiovisual, os dados contidos poderiam contribuir, por exemplo, subsidiando a escolha de um termo específico que represente assuntos relevantes de um determinado filme.

Este artigo utiliza uma fonte de dados (*Internet Movie Database* (IMDb)[4]) sobre filmes de um gênero específico (filmes de terror lançados entre 1980 e 1989) para realizar uma comparação entre dois conjuntos de dados. O uso do gênero terror justifica-se por sua definição a partir de elementos recorrentes, como mortos-vivos, bruxas e violência (Kawin, 2012). Por apresentar convenções temáticas e iconográficas marcadas, como criaturas, ambientes e tipos de ameaça, o gênero oferece um parâmetro adequado para verificar se as *tags* atribuídas pelos usuários captam essas convenções e em que medida as sinopses as expressam. A escolha do intervalo de tempo entre 1980 e 1989 se constitui como intencional e foi adotada para observar *tags* e sinopses atribuídas a filmes de um mesmo momento histórico do terror. Por serem filmes lançados há mais de três décadas, a pesquisa considera que esses filmes tiveram um longo período de exposição à comunidade de usuários da fonte dos dados.

O primeiro conjunto de dados foi coletado em um processo automatizado de mineração das palavras mais frequentes nos resumos de

[4] *Site:* https://www.imdb.com.

enredo (sinopses) enviados pelos usuários da plataforma. O segundo são termos de enredo (*plot keywords*) atribuídos pelos usuários. Essas palavras-chave funcionam como etiquetas, ou *tags*, por meio das quais os usuários buscam descrever aspectos do filme, como o termo *haunted house* (casa mal-assombrada), atribuído a *O Iluminado* (*The Shining*), de 1980. Os sistemas de atribuição de *tags*, contudo, oferecem a mecânica de atribuição, mas não regras sobre a semântica dos termos, de modo que as *tags* não correspondem necessariamente ao assunto do documento (Rafferty, 2018). Por isso, a pertinência dessas palavras-chave em relação ao enredo dos filmes constitui um dos aspectos analisados nesta pesquisa.

Comparar esses dados exige a utilização de abordagens interdisciplinares entre a representação temática dentro da Ciência da Informação e os estudos sobre filmes. Os estudos sobre gêneros cinematográficos mobilizam o terror e seus elementos mais comuns como um gênero popular. Para a pesquisa, os autores centrais nesse campo são Bruce F. Kawin (2012) e Julie Selbo (2015). O primeiro autor traz um panorama geral sobre o terror como gênero desde o seu nascimento até as suas características mais essenciais enquanto a segunda autora traz uma abordagem mais centrada nos gêneros cinematográficos como categorias narrativas, sendo o terror parte deste universo. A contribuição da Ciência da Informação reside no campo das folksonomias ou indexação social, como o verbete *tagging* da International Society for Knowledge Organization (ISKO) *Encyclopedia of Knowledge Organization* (Rafferty, 2018).

A intersecção entre esses autores e áreas pretende, junto à própria coleta de dados, oferecer caminhos para verificar em que medida as *tags* representam o conteúdo fílmico, em comparação com os termos mais frequentes extraídos das sinopses. Surge, assim, a questão norteadora da pesquisa: em que medida as *plot keywords* atribuídas pelos usuários do IMDb representam o conteúdo temático de filmes de terror lançados entre 1980 e 1989, em comparação com os termos mais frequentes extraídos de suas sinopses?

DOI http://dx.doi.org/10.1590/1981-5344/60611

2 PROCEDIMENTOS METODOLÓGICOS

A quantidade de bases de dados e informações relacionadas a cinema possibilita inúmeras fontes propícias a serem utilizadas no estudo. A escolha pelo IMDb, entretanto, se dá pela utilização recorrente desta base em pesquisas sobre filmes e dados, como indicado por Canet, Valero e Codina (2016). Soma-se ao exposto, a existência de um recurso denominado *storyline*, onde os dados de enredo estão agrupados. Dentro do campo *storyline*, o *website* oferece a possibilidade de os espectadores incluírem seus próprios termos de enredo – *plot keywords* em um sistema de indexação social (*tagging*) por relevância (termos mais bem avaliados como representativos aparecem em primeiro lugar). Esse sistema de relevância é baseado em *likes* e *dislikes* dos usuários. Os termos com mais *likes* são os mais bem avaliados. As *plot keywords* apresentam uma riqueza temática possível de ser utilizada para as pesquisas de representação da informação audiovisual.

Apesar da atribuição de *keywords* ser livre, o IMDb estabelece alguns critérios relacionados à forma de inclusão dessas *keywords*, como o uso de minúsculas, separação de palavras com hífen e verificação de termos já existentes. Em relação à pertinência, o portal indica que as *keywords* cadastradas relacionadas a subgêneros devem "apresentar papel significativo no enredo, ambiente ou meio do filme como um todo" (IMDb, 2026, tradução nossa), ou seja, o subgênero só pode ser atribuído a um filme se o elemento for central no filme, e não um detalhe. Um exemplo seria o uso inválido do termo "*jungle-adventure*" (aventura na selva) em filmes que possuem apenas uma única cena em uma selva. Para as demais *keywords*, não há exigência de que o termo represente uma característica central do enredo. A relevância de cada termo é estimada pelos votos dos usuários, o que indica validação social e não pertinência temática. Ressalta-se que as diretrizes consultadas correspondem à versão de 2026, posterior à coleta dos dados, realizada em 2024.

Os dois conjuntos de dados foram coletados por procedimentos distintos. As sinopses foram coletadas por meio da biblioteca *Cinemagoer* do *Python*, com uma busca por títulos que combinou o termo *horror* a cada ano entre 1980 e 1989. Dos resultados, foram mantidos os títulos lançados no ano correspondente que tivessem o terror (*horror*) entre os gêneros indexados e cujos resumos de enredo (*plot summaries*), enviados pelos usuários do IMDb, tivessem

CC BY

até 500 palavras, de modo a excluir textos excessivamente longos. Foram obtidos 127 títulos, cujos resumos foram reunidos e processados com a biblioteca *spaCy*[5]: os textos foram convertidos em minúsculas e tokenizados, as *stopwords* e os sinais de pontuação foram removidos e calculou-se a frequência das palavras, gerando uma lista com as 100 mais frequentes. O *script* utilizado pode ser acessado publicamente no *GitHub*[6]. As *plot keywords*, por sua vez, foram coletadas manualmente a partir da lista dos filmes do gênero terror mais populares lançados entre 1980 e 1989 [7], conforme a ordenação por popularidade da busca avançada do IMDb, considerando-se apenas o termo com maior número de votos de cada filme, até o total de 100 termos. Como os dois conjuntos derivam de listas de filmes distintas, a comparação foi realizada no nível do conjunto, e não filme a filme. Essas *keywords* foram agrupadas com base na distinção proposta por Kipp (2007) entre *tags* relacionadas ao assunto do documento e *tags* não relacionadas ao assunto, que não descrevem aquilo de que o documento (filme) trata.

Figura 1 - Pipeline ou fluxo de processos da pesquisa



Fonte: elaborado pelos autores (2024).

Por fim, as 100 *plot keywords* e as 100 palavras mais frequentes nas sinopses foram comparadas a fim de verificar em que medida representam o conteúdo temático dos filmes. Os termos dos dois conjuntos foram distribuídos pelo pesquisador nas seis dimensões de análise definidas na Seção 3.1: agentes de ameaça; acontecimentos; ambientação e comunidade; violência e corpo; sobrenatural e onírico; e subgêneros e franquias. Para o tratamento das diferenças lexicais, as formas flexionadas e derivadas de uma mesma palavra

[5] *Site*: https://spacy.io.
[6] *Site*: https://github.com/rafaelrbcarmo/ArtigoIMDB/blob/main/ScriptArtigoIMDBFinal.ipynb.
[7] *Site*: https://www.imdb.com/search/title/?release_date=1980-01-01,1989-12-31&genres=horror.

DOI http://dx.doi.org/10.1590/1981-5344/60611

foram agrupadas para a identificação das correspondências, como *kill*, *killed* e *killing*, *murder* e *murdered* ou *student* e *students*. Quanto ao grau de especificidade, distinguiram-se os termos específicos, que nomeiam um elemento particular do filme, como *space vampire* ou *killer car*, dos termos genéricos, que designam categorias amplas, como *monster* ou *man*. Consideraram-se correspondências os casos em que um mesmo termo, ou termos de mesma raiz, aparecem nos dois conjuntos dentro de uma mesma dimensão, como *werewolf* e *werewolf* ou *serial killer* e *killer*. Foram tratados como ruído e mantidos fora das dimensões os termos não relacionados ao assunto (Kipp, 2007), como *title directed by female*, os nomes próprios de personagens e os termos decorrentes de falhas da coleta, como *carvalho*, *rio*, *janeiro* e *brazil*. Também foram tratadas como ruído as palavras das sinopses que constituem metarreferências ao meio, e não ao conteúdo dos filmes, como *horror*, *movie*, *film*, [Televisão] *TV*, *channel*, *series*, *host* e *hosted*.

O referencial teórico caracteriza-se como uma revisão narrativa da literatura. Foram selecionadas obras dos estudos de cinema sobre gênero cinematográfico e o gênero terror e, no âmbito da Ciência da Informação, trabalhos sobre a representação da informação audiovisual e sobre a indexação social e as folksonomias. As obras foram identificadas em bases de dados como a Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI) e o Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), além das referências citadas nos próprios trabalhos recuperados. Adotaram-se como critérios de seleção a pertinência ao tema, a publicação em periódicos revisados por pares, livros ou eventos científicos e o idioma português ou inglês.

3 REVISÃO DE LITERATURA

Na história do cinema, padrões específicos de enredo evoluíram para estruturas temáticas mais definidas, mas ainda com certa flexibilidade. Essas estruturas são conhecidas como gêneros cinematográficos. A formação desse gênero como universo temático baseia-se principalmente em um conjunto de características comuns. Por exemplo, um filme ambientado em paisagens desertas do Oeste americano é identificado com o gênero faroeste (*western*). Já filmes que incluem monstros, fantasmas e elementos sobrenaturais tendem a

CC BY

ser associados ao terror (ou horror). Selbo (2015) denomina esse fenômeno como características ativas (no sentido em que oferecem pistas de sua intencionalidade de maneira mais clara ao argumentista) e passivas (que sugerem uma ideia central e outra imediatamente próxima, mas distinta) dos gêneros.

Os teóricos, por exemplo, questionam a forma como essa categorização é elaborada, os critérios utilizados e a tendência de se basearem em interesses de mercado (Correia, 2022). Destaca-se que a falta de precisão de uma categorização não se dá, segundo Correia (2022), pela falta de interesse de quem produz essas divisões (seja um fã, um crítico ou qualquer tipo de analista) e, sim, pela natureza complexa da própria ideia de categorias (ou gêneros).

Dado que os critérios que definem o gênero cinematográfico não são unânimes, é essencial que o organizador estabeleça, dentro dos estilos mais reconhecidos, meios de facilitar a localização de um filme. A criação de categorias ou gêneros muito distintos pode gerar lacunas e dificultar a visibilidade de um filme em meio a uma grande quantidade de outros.

Para fins de análise, a divisão dos gêneros mais reconhecidos proposta por Selbo (2015) se constitui como uma das mais sólidas dentro da literatura. A autora apresenta 31 gêneros possíveis – ação, aventura, terror, paródia, romance, melodrama, biografia, musical, mistério, crime, época, comédia, viagem, desastre, drama, sátira, dramédia, ficção científica, época, esporte, fantasia, super-herói, *noir*, *thriller*, guerra, histórico, faroeste e filmes em formatos específicos. Para o desenvolvimento do artigo, não será necessário analisar cada um destes gêneros. Se torna necessário, contudo, analisar o terror enquanto objeto de pesquisa.

3.1 O gênero terror: elementos temáticos principais

Selbo (2015, p. 59, tradução nossa) define o terror como:

> […] enredos em que forças maléficas (acontecimentos ou personagens) se infiltram ou assaltam o mundo cotidiano e perturbam a estabilidade da comunidade ou ordem mundial.

Retomando a definição de Kawin (2012, p. 4, tradução nossa), um filme de terror,

> […] é definido por seus elementos recorrentes (como mortos-vivos, bruxas ou violência grosseira e sangrenta), por suas atitudes em relação

DOI http://dx.doi.org/10.1590/1981-5344/60611

> a esses elementos (como que transgredir limites é perigoso) e por seu objetivo: assustar e revoltar o público.

Somam-se a essa definição, a delimitação dos subgêneros dentro deste universo.

O debate sobre subgêneros do terror parte do princípio das relações entre os conceitos gerais com as hibridizações possíveis entre esse gênero e outros, além da transformação das tendências de época em tipos a serem explorados. Alguns subgêneros relacionados a monstros (como vampiros e criaturas pequenas) refletem filmes que alcançaram sucesso comercial, como o subgênero *Aliens*, que surgiu na esteira de produções consagradas como *Alien* e *Predator*. Outros subgêneros como t*echnology*, *horror romance*, *gothic horror* e *comedy horror* indicam a fluidez do terror entre outros gêneros populares no cinema. A presença de subgêneros é evidenciada por autores como Massarolo e Gomes (2013) que evidenciam a presença de um vasto cânone dos filmes de zumbis em diferentes momentos históricos. Ainda que seja possível perceber como as tendências vão moldando o terror ao longo de sua história, não há, na literatura, um consenso sobre um critério único para definir os subgêneros do terror.

Existem classificações de gênero mais abrangentes e diretas (ou temáticas). O primeiro tipo traz, por exemplo, variações regionais (filmes de terror de um determinado país ou região), além de um foco maior em tendências relevantes do gênero como categorias a serem exploradas. O segundo tipo é mais voltado à tematização direta, onde os autores utilizam elementos mais comuns dentro dos filmes para elaborar uma classificação. São duas categorizações que guardam similaridades e se complementam.

A partir das definições apresentadas, é possível identificar os atributos que caracterizam tematicamente o filme de terror e que orientam, nesta pesquisa, a comparação entre os dois conjuntos de dados. A definição de Selbo (2015) aponta as forças maléficas, que podem assumir a forma de personagens ou de acontecimentos, e o mundo cotidiano e a comunidade cuja estabilidade é perturbada. Kawin (2012) acrescenta, entre os elementos recorrentes do gênero, a violência, enquanto a associação do terror a monstros, fantasmas e elementos sobrenaturais completa esse conjunto. Desses atributos derivam as dimensões de análise adotadas: (1) agentes de ameaça, que reúnem criaturas, monstros e assassinos; (2) acontecimentos, que reúnem os eventos que

CC BY

constituem a ameaça; (3) ambientação e comunidade, que reúnem os espaços, o tempo e os personagens afetados; (4) violência e corpo; e (5) sobrenatural e onírico. Os subgêneros e as franquias, discutidos anteriormente, foram considerados uma sexta dimensão, por reunirem esses atributos em rótulos reconhecidos pelos usuários e pela própria plataforma.

3.2 Folksonomia, IMDb e estudos com dados colaborativos sobre filmes

O relacionamento entre as categorias utilizadas na representação de filmes visando evitar ambiguidades é uma questão diretamente relacionada aos sistemas de recuperação enquanto ferramentas interativas entre os filmes e os espectadores. Os desafios referentes à navegabilidade, à localização de itens e à relevância de um determinado conteúdo integram o conjunto fundamental de análise da arquitetura da informação (Kipp, 2007, p. 2, tradução nossa).

A indexação social (ou folksonomia) preconiza a criação de "etiquetas (*tagging*) pelos usuários aos itens compartilhados" (Guedes; Dias, 2010, p. 47). Outra definição de indexação social foi apresentada por Rafferty (2018, tradução nossa) como "a prática em que usuários da internet geram *keywords* para descrever, categorizar ou comentar no conteúdo digital". Ao analisar os benefícios da folksonomia, conclui-se que ela é vantajosa pela liberdade de uso e pela facilidade de acesso, fatores que atraem o usuário e rapidamente o tornam parte do processo (Moraes; Lobo, 2020).

Rafferty (2018) compreende a atribuição de *tags* como uma forma de indexação, cujo valor se manifesta quando as *tags* de uma coleção são agregadas e compartilhadas em uma folksonomia. A autora ressalta que os sistemas de *tags* oferecem a mecânica de atribuição, mas não regras ou diretrizes sobre a semântica dos termos, de modo que os usuários decidem livremente a finalidade de suas *tags*, que não são necessariamente informativas ou relacionadas ao assunto do documento.

Como desvantagem, destaca-se um "número de informações que só podem ser recuperados por um pequeno grupo, uma vez que estes se utilizam de vocabulário próprio" (Moraes; Lobo, 2020, p. 122). Em bases de dados sobre o cinema, como o IMDb, é possível observar essa desvantagem nas *keywords*, que são termos categorizados por espectadores que devem versar sobre o

DOI http://dx.doi.org/10.1590/1981-5344/60611

conteúdo do filme. Contudo, a liberdade de inclusão permite que usuários registrem *plot keywords* que não possuem valor temático e essas *keywords* são misturadas às de valor temático interessante em termos de representação. Diante dessas limitações, a literatura discute formas de disciplinamento da atribuição de *tags*, como sistemas de recomendação de termos e o uso de vocabulários controlados. As diretrizes de atribuição de palavras-chave do IMDb, apresentadas na Seção 2, podem ser compreendidas como uma forma desse disciplinamento (Rafferty, 2018). Rafferty (2018) conclui que as *tags* tendem a complementar, e não a substituir, os sistemas convencionais de organização do conhecimento, oferecendo oportunidades para a indexação do conteúdo e das respostas emocionais, sobretudo em recursos não textuais, como imagens e filmes.

A utilização dos gêneros e subgêneros cinematográficos enquanto dados é um fenômeno relativamente recente, estimulada, principalmente, pelo surgimento das bases de dados na Internet. Um dos exemplos mais relevantes de centro de dados sobre o audiovisual é o próprio IMDb. Essa plataforma, financiada e mantida pela Amazon, possui uma quantidade considerável de informações sobre filmes, indo desde as informações técnicas quanto temáticas. O IMDb ganhou uma importância relevante na percepção dos usuários sobre filmes, principalmente ao se considerar as interações possíveis para os usuários dentro da plataforma, como a criação de uma conta pessoal, onde é possível indicar quais filmes gostaria de assistir, indicar os já vistos, adicionar análises ou comentários.

O IMDb também funciona como uma *wiki* especializada em cinema, onde os usuários podem editar as informações e inserir tópicos (validados por outros usuários). A opção de editar permite inserir *plot keywords*, palavras-chave que sintetizam, de maneira resumida, o gênero e outros termos que possam representá-lo. A plataforma, nesse sentido, se adequa ao status de objeto de estudo da computação social, que "pode ser também usada como um termo abrangente pra toda pesquisa de ciência da computação que analisa o conteúdo e a atividade em redes sociais" (Manovich, 2015, p. 69). Topal e Ozsoyoglu (2016, p. 1170) ratificam a importância desta base, indicando que o IMDb se constitui como a "fonte de conteúdo para filmes, TV e celebridades

CC BY

mais popular do mundo". A utilização maciça da plataforma a transformou em uma ferramenta útil para pesquisadores e curiosos sobre o cinema.

Kipp (2007, p. 1, tradução nossa) enumera autores que argumentam, ao apresentar *tags* de classificações por frequência ou em formato de nuvem, que "os sistemas de etiquetas formam novas taxonomias ou folksonomias interessantes de termos relacionados". A pesquisadora afirma que, embora algumas etiquetas cadastradas possam ser aproveitadas como termos de indexação, muitas outras *tags* são afetivas, que residem na experiência do usuário e, nas palavras da autora, de uma resposta emocional, ou são relacionadas a marcações de tarefas relacionadas a essas obras. Como exemplo, as *tags* afetivas coletadas no estudo expressavam entusiasmo ou emoção. Já as *tags* de tempo ou tarefa se constituíam em *toread* ("para ler"), *tobuy* ("para comprar") e suas variantes ortográficas. A autora indica que a análise dessas etiquetas afetivas e de tempo e tarefas pode indicar como os usuários classificam e organizam a informação (Kipp, 2007, p. 3). Outro exemplo dessa tendência é verificado em uma pesquisa sobre as *tags* de um *site* de música.

> [...] Encontramos, por exemplo, que os usuários do Last.fm replicam o comportamento de outros usuários com menos frequência do que pesquisas anteriores em outros sistemas de marcação sugerem. Este estudo também é o primeiro a examinar a dinâmica da marcação social de música. Descobrimos que, no final da primeira semana após o seu lançamento, a maioria das músicas não foi marcada e que uma grande proporção das primeiras marcações aplicadas a uma música são relacionadas a gênero e de natureza ampla. Finalmente, observamos uma proporção significativa de marcações extensas, às vezes contendo até 14 palavras, muitas das quais são poéticas e muito pessoais. Isso sugere que, para alguns usuários, as marcações sociais são usadas como um meio de expressão de sua relação com a música e, até certo ponto, de sua própria identidade (Laplante, 2015, p. 52, tradução nossa).

Kipp (2007) observa, em sua pesquisa, a presença de preposições nas etiquetas sugeridas por usuários dessas plataformas, em vez de palavras individuais ou compostas significativas. Esse fato indica que alguns usuários podem não estar totalmente cientes de como os vários sites formam etiquetas. Kipp (2007) conclui que há uma tendência entre os usuários de atribuir etiquetas a partir de uma perspectiva holística, considerando não só o conteúdo, mas também elementos de sua própria individualidade como relevantes para a recuperação da informação (Kipp, 2007).

DOI http://dx.doi.org/10.1590/1981-5344/60611

A aplicação da mineração textual em relação a *tags* relacionadas a cinema foi explorada em diversos contextos. Srivastava, Singh e Tapaswi (2023) constroem uma ferramenta de análise emocional do enredo de filmes nos dados disponíveis no *MovieGraphs*[8]. O trabalho dos autores é uma importante contribuição para a pesquisa na medida em que criam um recurso – neste caso, a ferramenta de análise emocional do enredo – por meio de dados disponíveis de forma colaborativa em uma base de informações mais ampla. Sirisha, Devi e Ramesh (2022) por sua vez, trabalham diretamente com elementos de enredo de filmes (*plot-topic*) utilizando a modelagem de tópicos. Os autores verificaram que a recomendação de filmes de um mesmo gênero, a partir da rotulagem padrão, não vem atraindo mais os espectadores. A pesquisa trabalha com a construção de uma modelagem de tópicos para atribuir uma rotulagem baseada no enredo de cada filme. Os dados utilizados foram minerados a partir de listas dos espectadores feitas no IMDb. As listas contêm informações como sinopse, elementos do enredo e outros aspectos colaborativos dentro da plataforma. O processamento dos dados da pesquisa também foi inspirado neste trabalho, pois os autores utilizam a ferramenta *WordNet*[9].

Dhir e Raj (2018) e Darapaneni *et al*. (2020) desenvolveram diferentes algoritmos para prever ou não o sucesso de um filme antes de seu lançamento, também utilizando os dados do IMDb. Os dois trabalhos são relevantes no sentido que realizam aplicações dos gêneros cinematográficos para prever tendências de mercado. Çizmeci e Öğüdücü (2018) por sua vez, utilizam abordagens particulares para desenvolver mecanismos de previsão do sucesso de um filme.

4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS OBTIDOS

A coleta manual das *plot keywords* foi realizada a partir da lista dos filmes mais populares do IMDb e lançados entre 1980 e 1989. A opção por esse intervalo se deu para verificar tendências nos filmes do período. Para a pesquisa

[8] O *MovieGraphs* é um conjunto de dados criado para auxiliar na compreensão de situações sociais humanas em vídeos. Foi desenvolvido por pesquisadores para estudar interações sociais e contextos emocionais a partir de cenas de filmes.
[9] O *WordNet* é uma base de dados lexical em inglês que organiza palavras em grupos de sinônimos chamados *synsets* (conjuntos de sinônimos). Cada *synset* representa um conceito específico e inclui palavras que compartilham o mesmo significado dentro de determinados contextos.

CC BY

foi coletado apenas o termo com a maior quantidade de *likes* de cada filme da lista.

Quadro 1 – Lista das 100 *plot keywords* coletadas manualmente no IMDb.

100 *plot keywords* mais bem avaliadas de filmes de terror dos anos 80
haunted house (18),actual animal killed (8), evil dead (7), isolation(7), vampire(7), florist(7), reanimation(7), christmas(6),slasher (6),werewolf(6),space vampire(6), giant metal phallus(6), sadomasochism(5), surprise ending(5), intentionally misspelled title(5), surrealism(5), hospital(5), jungle(4),alien(4), poltergeist(4), semi truck driving itself(4), psychic power(4), title directed by female (4), post apocalypse(4), hallucination(4), fog(4), based on novel(4), demonic possession(4), sex scene (3),corn(3), horror host (3), detective(3), maniac(3), fake identity(3), mansion (3), coma(3), inheritance(3),based on the works of Stephen King(3), dream(3), cannibal tribe(3), alligator(3), carnival(3), gay(3), bully(3), isolation tank(3), body horror(3),nightmare(3), killer car(3), creature feature(3), grandmother(3),summer camp(3), identical twins(3), serial killer(3), rabies(3), living dead(3), female full frontal nudity(3), sexuality(3), zombie invasion(3), summer (2), teleportation (2), anthology(2), pitchfork(2), andy barclay character(2), 1980s(2), friday the thirteenth(2), maggot(2), puppet show(2), mirror(2), truck(2), eaten alive(2), reference to schrodinger's cat(2), sex in bed(2), monster (2), female rear nudity(2), extreme film(2), dildo(2), absurd violence(2), swamp(2), pantyhose(2), vietnam war veteran(2), insane asylum(2), directorial debut(2), writer(2), american native(2), cientific(2), hell(2), racist(2), grindhouse film(1), slasher(1),jason voorhees Character(1), dark comedy(1), female stockinged legs(1), supernatural serial murder(1), good versus evil (1), , divorced woman(1), armed robbery(1), murder(1), , cult film(1), survival (1), dog(1), sheep (1)

Fonte: elaborado pelos autores (2024).

O termo com a maior relevância (quantidade de *likes*) foi *haunted house*, coletado do filme mais influente do período, O Iluminado (*The Shining*). Em um segundo grupo, destacam-se termos com oito, sete e seis *likes*. O mais relevante neste segundo grupo (*actual animal killed*) está relacionado a acontecimento real.

O grupo entre cinco e três *likes* apresenta muitas *plot keywords* que podem ser associadas a aspectos técnicos ou convenções presentes em filmes (*surprise ending*, *intentionally misspelled title*, *title directed by female*, *based on novel*, *based on the works of Stephen King*). Esses termos correspondem ao que Kipp (2007) denomina *tags* não relacionadas ao assunto, pois não descrevem o enredo dos filmes, e alguns deles figuram entre os mais votados, como *actual animal killed*. Esse resultado evidencia que a validação social por meio de votos não garante a pertinência temática do termo. Outro conjunto desses termos pode estar relacionado a ambientação, locais ou contexto desses filmes (*hospital*, *jungle*, *mansion*, *isolation tank*, *summer camp*, *zombie invasion*, *fog*, *semi truck driving itself*, *corn*, *carnival*, *post apocalypse*). Percebe-se também um grupo de *plot keywords* de nome de criaturas, tipos ou personagens icônicos (*maniac ,alien*, *alligator*, *killer car*, *creature feature*, *living dead*, *serial killer*,

DOI http://dx.doi.org/10.1590/1981-5344/60611

cannibal tribe, *poltergeist*, *horror host*, *grandmother*, *identical twins*, *detective*, *fake identity*). Outro padrão está relacionado a termos sexuais e gêneros (*sadomasochism*, *sex scene*, *female full frontal nudity*, *sexuality*, *gay*).

Ainda é possível verificar *plot keywords* que representam aspectos oníricos ou simbólicos (*surrealism, hallucination, dream, psychic power, nightmare, demonic possession, coma, body horror*), relações de parentesco (*grandmother*, *inheritance*), abuso e doenças (*bully*, *rabies*). Por fim, as *plot keywords* que receberam um ou dois *likes* possuem temas similares aos encontrados nos grupos anteriores.

Quadro 2 – 100 palavras mais frequentes em sinopses coletadas via script

100 palavras mais frequentes em sinopses de filmes dos anos 80
'horror': 94, 'house': 50, 'movie': 47, 'evil': 42, 'new': 41, 'group': 35, 'night': 34, 'town': 33, 'professor': 32, 'young': 30, 'TV': 30, 'years': 27, 'people': 27, 'hosted': 26, 'father': 26, 'soon': 25, 'find': 25, 'man': 25, 'film': 23, 'mother': 23, 'city': 22, 'channel': 22, 'friends': 22, 'killer': 22, 'life': 22, 'dead': 21, 'billy': 21, 'killed': 20, 'family': 20, 'local': 20, 'murder': 20, 'eddie': 20, 'story': 19, 'old': 19, 'stop': 19, 'death': 18, 'mysterious': 18, 'de': 18, 'begins': 18, 'finds': 17, 'movies': 17, 'lake': 17, 'boy': 17, 'killing': 17, 'host': 17, 'werewolf': 17, 'brother': 16, 'carvalho': 16, 'rio': 16, 'janeiro': 16, 'brazil': 16, 'monster': 16, 'girl': 16, 'woman': 16, 'hospital': 16, 'count': 16, 'late': 16, 'time': 16, 'rock': 16, 'little': 15, 'gets': 15, 'blood': 15, 'dr': 15, 'son': 15, 'goes': 15, 'students': 15, 'store': 15, 'plant': 15, 'audrey': 15, 'friend': 14, 'murdered': 14, 'police': 14, 'deadly': 14, 'high': 14, 'santa': 14, 'seymour': 14, 'paul': 14, 'book': 13, 'mark': 13, 'discovers': 13, 'world': 13, 'tom': 13, 'later': 13, 'children': 13, 'help': 13, 'remote': 13, 'home': 13, 'tina': 13, 'series': 12, 'haunted': 12, 'student': 12, 'owner': 12, 'tribe': 12, 'returns': 12, 'daughter': 12, 'day': 12, 'cabin': 12, 'small': 12, 'college': 12, 'mansion': 12.

Fonte: elaborado pelos autores (2024).

A palavra de maior frequência, 'horror' (94), tem sua incidência ampliada pela própria estratégia de busca, que utilizou esse termo, e funciona como uma metarreferência ao gênero. Palavras como '*movie*' (47) e '*film*' (23), juntamente com '*TV*' (30), '*channel*' (22), '*hosted*' (26), '*host*' (17) e '*series*' (12), sugerem uma forte presença de narrativas que se desenrolam no universo da mídia e do cinema, o que se explica pela estratégia de busca, que recuperou títulos com o termo *horror*, entre os quais programas de televisão apresentados por anfitriões de terror e compilações sobre o gênero. Por constituírem metarreferências ao meio, e não ao conteúdo dos filmes, essas palavras foram tratadas como ruído na comparação.

A incidência de palavras que denotam locais e ambientes é notável, com '*house*' (50) se destacando, confirmando a casa assombrada ou o lar invadido como um cenário primordial no terror oitentista. '*town*' (33) e '*city*' (22) indicam o contraste entre ambientes urbanos e pequenos centros, enquanto '*night*' (34) estabelece a atmosfera temporal.

Em relação aos personagens e grupos, '*group*' (35), '*people*' (27), '*young*' (30) e '*man*' (25) são termos genéricos de alta frequência, indicando a presença de múltiplos indivíduos, muitas vezes jovens, como protagonistas ou vítimas. A estrutura familiar também é proeminente com '*father*' (26), '*mother*' (23), '*family*' (20), '*boy*' (17), '*brother*' (16), '*girl*' (16), '*woman*' (16), '*son*' (15), '*children*' (13) e '*daughter*' (12), evidenciando o terror se infiltrando no núcleo familiar. Termos como '*friends*' (22), '*professor*' (32), '*students*' (15) e '*student*' (12) sugerem cenários escolares.

Outros temas são representados por palavras como '*evil*' (42), '*killer*' (22), '*dead*' (21), '*killed*' (20), '*murder*' (20), '*death*' (18) e '*killing*' (17), sublinhando a presença constante de violência, assassinato e morte. A atmosfera de mistério é evocada por '*mysterious*' (18). Criaturas sobrenaturais ou monstruosas aparecem com '*werewolf*' (17) e '*monster*' (16), enquanto '*blood*' (15) remete a um elemento comum em filmes gráficos ou violentos. Termos como '*haunted*' (12) e '*deadly*' (14) contribuem para a caracterização de locais e ameaças.

A lista também revela a incidência de verbos de ação que impulsionam a narrativa, como '*find*' (25), '*soon*' (25), '*stop*' (19), '*begins*' (18), '*finds*' (17), '*goes*' (15), '*discovers*' (13) e '*returns*' (12), indicando a progressão do enredo em torno de descobertas, inícios de eventos e retornos. A presença das palavras como '*new*' (41) e '*old*' (19) pode refletir a introdução de novas ameaças ou o retorno de males antigos. A palavra '*world*' (13) sugere a escala global ou a ameaça a toda a existência.

Mesmo incluindo filtros, algumas limitações do *Cinemagoer* ainda ocasionaram palavras vazias ou distorcidas. Ao verificar a incidência de 'carvalho' (16), 'rio' (16), 'janeiro' (16) e 'brazil' (16), constatou-se que se referiam à identificação do autor de um dos resumos de enredo, incluída nos dados coletados. Apesar deste tipo de distorção, foi um fenômeno isolado que não invalida as outras palavras coletadas. Palavras como '*billy*' (21), '*eddie*' (20), '*audrey*' (15), '*seymour*' (14), '*paul*' (14), '*mark*' (13) e '*tom*' (13), '*tina*' (13), são nomes próprios que aparecem com frequência notável, sugerindo personagens recorrentes ou títulos de filmes específicos que foram proeminentes na amostra.

DOI http://dx.doi.org/10.1590/1981-5344/60611

Quadro 3 - Distribuição das *keywords* e das palavras das sinopses nas dimensões de análise

Dimensão	*Keywords* (nº)	Exemplos	Sinopses (nº)	Exemplos	Correspondências
Agentes de ameaça	15	*Space vampire, killer car, serial killer, poltergeist, cannibal tribe*	8	evil, killer, werewolf, monster, plant, tribe	*werewolf*, *monster*; *killer / serial killer*, *killer car*; *tribe / cannibal tribe*
Acontecimentos	7	*reanimation, teleportation, armed robbery, survival, fake identity*	0	-	-
Ambientação e comunidade	28	*Haunted house, summer camp, insane asylum, christmas, florist, vietnam war veteran*	33	house, town, city, lake, night, family, father, young, students	*mansion, hospital; house / haunted house*
Violência e corpo	6	*eaten alive, maggot, pitchfork, absurd violence, rabies*	8	killed, murder, death, dead, blood	*murder*; *dead / living dead*
Sobrenatural e onírico	9	*demonic possession, psychic power, dream, nightmare, hell*	2	haunted, mysterious	*haunted / haunted house*
Subgêneros e franquias	13	*slasher, body horror, creature feature, zombie invasion, friday the thirteenth*	0	-	-
Temas não previstos nas dimensões	11	*sex scene, sexuality, female full frontal nudity*	0	-	-
Não temáticos ou ruído	11	*title directed by female, based on novel, horror host, cult film*	24	metarreferências (horror, movie, TV, host) e nomes próprios (billy, eddie, carvalho, rio)	-
Genéricos sem dimensão	0	-	25	new, soon, find, begins, returns, years, life	-
Total	100	-	100	-	-

Fonte: elaborado pelos autores (2024).

CC BY

A distribuição dos termos nas dimensões de análise (Quadro 3) evidencia diferenças de cobertura e de especificidade entre os dois conjuntos. As *plot keywords* contemplam as seis dimensões, com predomínio de ambientação e comunidade (28) e de agentes de ameaça (15), e nomeiam elementos particulares dos filmes, como *space vampire*, *killer car* e *summer camp*. As palavras das sinopses concentram-se em ambientação e comunidade (33), com termos genéricos como *house*, *town* e *family*, e não contemplam as dimensões de acontecimentos e de subgêneros e franquias; além disso, 49 das 100 palavras são metarreferências, nomes próprios ou termos genéricos sem dimensão definida. As correspondências diretas entre os conjuntos são poucas e concentram-se nos agentes de ameaça e na ambientação, como *werewolf*, *monster*, *mansion* e *hospital*. Observa-se ainda que 11 *plot keywords* referem-se à sexualidade e à nudez, aspecto não previsto nas definições do gênero adotadas, e outras 11 não se relacionam ao assunto dos filmes.

Os dois conjuntos de dados apresentam relações entre si, embora, isoladamente, sua capacidade de representação do conteúdo fílmico seja limitada. As *plot keywords* concentram-se em elementos do enredo, como criaturas, personagens, eventos e ambientes, mas incluem também termos que, na distinção proposta por Kipp (2007), não se relacionam ao assunto do documento. Diferentemente do observado pela autora em sistemas de favoritos sociais, essas *tags* não são de tempo e tarefa, como *toread*, nem predominantemente afetivas: referem-se a aspectos de produção, à origem da obra ou a curiosidades, como *title directed by female* e *intentionally misspelled title*. Outra diferença diz respeito à forma dos termos: enquanto Kipp (2007) identificou *tags* formadas por frases com preposições e conjunções, as *plot keywords* analisadas consistem em palavras individuais ou compostas, como *haunted house* e *isolation*, o que pode ser atribuído às diretrizes de formato da plataforma (IMDb, 2026). Esses resultados confirmam a observação de Rafferty (2018) de que, na ausência de regras sobre a semântica dos termos, as *tags* não são necessariamente relacionadas ao assunto do documento, ainda que o disciplinamento da forma de registro produza termos mais padronizados.

As palavras mais frequentes das sinopses, por sua vez, apresentam sentidos predominantemente genéricos, o que se justifica, em parte, pela natureza narrativa desse tipo de texto e pela composição da amostra. Ainda

DOI http://dx.doi.org/10.1590/1981-5344/60611

assim, convergem com as *plot keywords* em elementos centrais do gênero, como *house* e *haunted*, *werewolf*, *monster* e *killer*. A diferença entre os dois conjuntos pode ser compreendida a partir da distinção apresentada por Rafferty (2018) entre os planos paradigmático e sintagmático da descrição: a atribuição de *tags* opera predominantemente no plano paradigmático, por meio de termos isolados ou frases curtas, enquanto descrições narrativas, como as sinopses, situam-se no plano sintagmático e permitiriam captar uma gama mais ampla de interpretações. Nesse sentido, palavras como *murder*, *killer* e *young*, frequentes nas sinopses, contribuem para contextualizar *plot keywords* como *slasher* ou *summer camp*, o que indica uma relação de complementaridade entre os dois conjuntos.

5 CONSIDERAÇÕES FINAIS

A utilização de dados para a categorização de filmes é explorada de maneira diversa na literatura, sobretudo em trabalhos que abordam a construção de modelos, como os que buscam prever o sucesso de um filme antes de seu lançamento. Esta pesquisa, por sua vez, buscou verificar em que medida as *plot keywords* atribuídas pelos usuários do IMDb representam o conteúdo temático de filmes de terror lançados entre 1980 e 1989, em comparação com os termos mais frequentes extraídos de suas sinopses. Com base nas dimensões de análise derivadas das definições do gênero, os resultados indicam que as *plot keywords* contemplam todas as dimensões consideradas (agentes de ameaça, acontecimentos, ambientação e comunidade, violência e corpo, sobrenatural e onírico, e subgêneros e franquias) e o fazem de forma mais específica do que as sinopses. As palavras mais frequentes das sinopses concentram-se em ambientação e comunidade, com termos predominantemente genéricos, e não contemplam as dimensões de acontecimentos nem de subgêneros e franquias. A representação oferecida pelas *plot keywords*, contudo, é parcial: parte delas refere-se a aspectos de produção, à origem da obra ou a curiosidades, e outras tratam de sexualidade e nudez, aspecto não previsto nas definições do gênero adotadas, o que evidencia que a validação social por meio de votos não garante a pertinência temática dos termos. As correspondências diretas entre os dois conjuntos são poucas e concentram-se nos agentes de ameaça e na ambientação, o que

CC BY

indica que operam em planos distintos de descrição e devem ser reconhecidos como formas complementares, e não substitutas, de representação do conteúdo fílmico. O debate sobre gêneros cinematográficos mostra-se, nesse contexto, importante para avaliar terminologicamente as *tags* utilizadas, oferecendo uma base teórica para enriquecer sistemas de classificação de filmes.

As limitações da pesquisa se deram principalmente em relação ao corpus de análise. Para um resultado mais efetivo, se torna necessário utilizar bibliotecas mais amplas, ainda que exista uma dificuldade de se extrair dados sobre filmes em bases como o IMDb. Nesse sentido, a limitação de alguns sites sobre quais dados podem ser extraídos foi outro fator a ser considerado. Em termos de coleta, a biblioteca *Cinemagoer* apresentou algumas falhas que impediram resultados mais definitivos, tornando a busca por outras bibliotecas mais eficientes uma necessidade. A dificuldade em encontrar análises mais específicas sobre os termos utilizados também tornou a pesquisa algo desafiador.

Acrescenta-se que o recorte de um único gênero e de uma única década não permite generalizar os resultados para outros períodos ou gêneros, e que as diretrizes da plataforma consultadas correspondem a uma versão posterior à coleta dos dados. Ressalta-se, ainda, que a distribuição dos termos nas dimensões de análise resulta de uma codificação interpretativa realizada pelo pesquisador, e que parte dos termos admite mais de uma classificação. A coleta das sinopses por meio de busca por título, e não pela lista de popularidade, resultou em uma amostra que inclui programas de televisão e compilações, e os dois conjuntos de dados não se referem integralmente aos mesmos filmes. A pesquisa se soma aos esforços de aproximar ciência da informação, análise de dados e a pesquisa em cinema. Trata-se de uma pesquisa inicial que pode ser expandida por novas perspectivas como a aplicação de uma categorização baseada nos gêneros cinematográficos em sistemas automáticos de recuperação.

DOI http://dx.doi.org/10.1590/1981-5344/60611

REFERÊNCIAS

CANET, Fernando; VALERO, Miguel Ángel; CODINA, Lluís. Quantitative approaches for evaluating the influence of films using the IMDb database. **Communication & Society**, v. 29, n. 2, p. 151-172, jan. 2016. Universidad de Navarra. DOI 10.15581/003.29.2.151-172.

CORDEIRO, Rosa Inês Novaes. Informação cinematográfica e textual: da geração à interpretação e representação de imagem e texto. **Ciência da Informação**, v. 25, n. 3, 1996. DOI 10.18225/ci.inf.v25i3.648.

CORREIA, Gabriel Costa. O gênero horror e o arqui-gênero: subgêneros e transtextualidade em A casa muda, Audition e a Centopéia humana. **Cambiassu**: Estudos em Comunicação, v. 7, n. 10, p. 31-40, jan. 2022 Disponível em: https://periodicoseletronicos.ufma.br/index.php/cambiassu/article/view/18509. Acesso em: 10 jul. 2025.

ÇIZMECI, Beyza; ÖĞÜDÜCÜ, Sule Gunduz. Predicting IMDb Ratings of Pre-release Movies with Factorization Machines Using Social Media. *In*: INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE AND ENGINEERING (UBMK), 3., 2018, Sarajevo. **Proceedings** […]. Sarajevo: IEEE, Sept. 2018. p. 173-178. DOI 10.1109/ubmk.2018.8566661.

DARAPANENI, Narayana; BELLARMINE, Christopher; PADURI, Anwesh Reddy; ENTOORI, Sujana; KUMAR, Abir; VYBHAV, S V; MONDAL, Koushik. Movie Success Prediction Using ML. *In*: IEEE ANNUAL UBIQUITOUS COMPUTING, ELECTRONICS & MOBILE COMMUNICATION CONFERENCE (UEMCON), 11., 2020, New York. **Proceedings** […]. New York: IEEE, Oct. 2020. p. 0869-0874. DOI http://dx.doi.org/10.1109/uemcon51285.2020.9298145.

DHIR, Rijul; RAJ, Anand. Movie Success Prediction using Machine Learning Algorithms and their Comparison. *In*: INTERNATIONAL CONFERENCE ON SECURE CYBER COMPUTING AND COMMUNICATION (ICSCCC), 1., 2018, Jalandhar. **Proceedings** […]. Jalandhar: IEEE, Dez. 2018. p. 385-390. DOI http://dx.doi.org/10.1109/icsccc.2018.8703320.

GUEDES, Roger de Miranda; DIAS, Eduardo José Wense. Indexação social: abordagem conceitual. **Revista ACB**: Biblioteconomia em Santa Catarina, v. 15, n. 1, 2010. Disponível em: https://revista.acbsc.org.br/racb/article/view/686. Acesso em: 10 jun. 2025.

IMDB. **Keywords**. [*S. l.*]: IMDb, 2026. Disponível em: https://help.imdb.com/article/contribution/titles/keywords/GXQ22G5Y72TH8MJ5. Acesso em: 18 out. 2026.

KAWIN, Bruce. **Horror and the horror film**. Colorado: Anthem Press, 2012.

KIPP, Margaret E. I. @toread and Cool: tagging for time, task and emotion. *In*: INFORMATION ARCHITECTURE SUMMIT, 8., 2007, Arizona. **Proceedings** [...]. Arizona: Information Architecture Summit, 2007. p. 1-11. Disponível em: https://repository.arizona.edu/bitstream/handle/10150/105606/mkipp-iasummit2007.pdf?sequence=1&isAllowed=y. Acesso em: 10 jun. 2025.

LAPLANTE, Audrey. Etiquetagem em primeira audição: uma análise das práticas de etiquetagem social em um sistema de recomendação de música. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 20, n. [especial], p. 33-54, 2015. DOI https://doi.org/10.5007/1518-2924.2015v20nesp1p33.

MANOVICH, Lev. A Ciência da Cultura? Computação Social, Humanidades Digitais e Analítica Cultural. **MATRIZes**, São Paulo, v. 9, n. 2, p. 67-84, 2015. DOI https://doi.org/10.11606/issn.1982-8160.v9i2p67-84.

MASSAROLO, Joao Carlos; GOMES, Paula. Um estudo sobre construção de mundos no cinema de terror: representações das multidões nos filmes de zumbi. **Revista Eco-Pós**, v. 15, n. 3, p. 196-216, 2013. DOI https://doi.org/10.29146/eco-pos.v15i3.879.

MORAES, Luana Barbosa; LOBO, Paulo Mauricio Santos. Folksonomia: a tagzação da informação na era digital. **Revista Bibliomar**, v. 19, n. 1, p. 110-124, jun. 2020. Disponível em: https://periodicoseletronicos.ufma.br/index.php/bibliomar/article/view/14216. Acesso em: 10 jul 2025.

RAFFERTY, Pauline. Tagging. *In*: HJØRLAND, Birger; GNOLI, Claudio (ed.). **ISKO Encyclopedia of Knowledge Organization**. [*S. l.*]: ISKO, 2018. Disponível em: https://www.isko.org/cyclo/tagging.htm. Acesso em: 3 ago. 2025.

SELBO, Julie. **Film genre for the screenwriter**. California: Taylor & Francis Group, 2015.

SIRISHA, Potnuru; DEVI, G. Lavanya; RAMESH, N. Plot-Topic based Movie Recommendation System using WordNet. *In*: INTERNATIONAL CONFERENCE ON COMPUTING FOR SUSTAINABLE GLOBAL DEVELOPMENT (INDIACOM), 9., 2022, New Delhi. **Proceedings** [...]. New Delhi: IEEE, Mar. 2022. p. 45-4. DOI 10.23919/indiacom54597.2022.9763244.

SRIVASTAVA, Dhruv; SINGH, Aditya Kumar; TAPASWI, Makarand. How you feelin'? Learning emotions and mental states in movie scenes. *In*: IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 1., 2023, Vancouver. **Proceedings** [...]. Piscataway: IEEE, 2023. p. 2517-2528. DOI 10.1109/CVPR52729.2023.00248.

TOPAL, Kamil; OZSOYOGLU, Gultekin. Movie review analysis: emotion analysis of imdb movie reviews. *In*: IEEE/ACM INTERNATIONAL CONFERENCE ON ADVANCES IN SOCIAL NETWORKS ANALYSIS AND MINING (ASONAM), 1., 2016, San Francisco. **Proceedings** [...]. San Francisco: IEEE, ago. 2016. p. 1170-1176. DOI 10.1109/asonam.2016.7752387.

DOI http://dx.doi.org/10.1590/1981-5344/60611

CONFLITOS DE INTERESSE

Os autores declaram não haver conflito de interesses de ordem pessoal, financeira, intelectual, profissional, política ou religiosa no desenvolvimento e na publicação deste artigo.

CC BY

CONTRIBUIÇÕES DAS AUTORIAS

Informa-se nesta seção as funções de cada autoria, de acordo com a taxonomia CRediT, conforme orientado na página da revista PCI:

Função	Definição
Conceituação	Rafael Rodrigo do Carmo Batista; Thiago Henrique Bragato Barros.
Curadoria de dados	Rafael Rodrigo do Carmo Batista.
Análise Formal	Rafael Rodrigo do Carmo Batista; Thiago Henrique Bragato Barros.
Obtenção de financiamento	—
Investigação	Rafael Rodrigo do Carmo Batista; Thiago Henrique Bragato Barros.
Metodologia	Rafael Rodrigo do Carmo Batista; Thiago Henrique Bragato Barros.
Administração do projeto	—
Recursos	—
Software	—
Supervisão	—
Validação	—
Visualização [de dados (infográfico, fluxograma, tabela, gráfico)]	—
Escrita – primeira redação	—
Escrita – revisão e edição	Rafael Rodrigo do Carmo Batista; Thiago Henrique Bragato Barros.