



HAL
open science

Recommander autrement les films documentaires ? Design d'une preuve de concept basée sur l'indexation des dispositifs de réalisation

Samuel Gantier, Eve Givois, Eric Kergosien

► To cite this version:

Samuel Gantier, Eve Givois, Eric Kergosien. Recommander autrement les films documentaires ? Design d'une preuve de concept basée sur l'indexation des dispositifs de réalisation. *Revue des Interactions Humaines Médiatisées (RIHM) = Journal of Human Mediated Interactions*, 2020, 21 (2), pp.1-26. hal-03466413

HAL Id: hal-03466413

<https://hal.science/hal-03466413v1>

Submitted on 1 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Recommander autrement les films documentaires ? Design d'une preuve de concept basée sur l'indexation des dispositifs de réalisation

Recommending documentary films differently? Design of a Proof of Concept based on the indexing of filmmaking practices

Samuel GANTIER (1), Ève GIVOIS (1), Éric KERGOSIEN (2)

(1) DeVisu, Université Polytechnique Hauts-de-France
samuel.gantier@uphf.fr ; eve.givois@uphf.fr

(2) GERiCO, Université de Lille
eric.kergosien@univ-lille.fr

Résumé. Cet article présente le processus de conception et d'évaluation d'une preuve de concept développée dans le cadre du programme AlgoDoc (Algorithme de recommandation de films documentaires). L'objectif est d'expérimenter une recommandation basée sur une nouvelle forme de catégorisation du cinéma documentaire alternative aux métadonnées fondées principalement sur une classification thématique. Le terrain expérimental se compose d'un catalogue d'environ 800 documentaires de création (proposé à un public cinéphile par la plateforme de vidéo à la demande Tënk) et du moteur algorithmique Rumo (mis à disposition par la société Spideo). Dans un premier temps, un corpus de 112 films a été indexé via un thésaurus créé spécifiquement pour cette étude dans l'optique de qualifier la grande variété des dispositifs de réalisation du cinéma documentaire. Dans un deuxième temps, le recueil des recommandations d'un panel d'experts du domaine a permis d'analyser les écarts avec la recommandation algorithmique. Les résultats mettent en avant de nouveaux réseaux de liens entre les films qui n'existaient pas au préalable avec la classification thématique. Ils permettent d'ébaucher de nouvelles perspectives pour améliorer la médiation numérique du cinéma documentaire sur les plateformes.

Mots-clés. Cinéma documentaire, plateforme de vidéo à la demande, indexation audiovisuelle, algorithme de recommandation, preuve de concept, design d'expérience utilisateur, médiation numérique.

Abstract. This article presents the evaluation of a proof of concept whose objective is to establish an algorithmic recommendation of documentary films based on the filmmaking practices. The challenge is to overcome the limits of an exclusively thematic indexing of documentary films. The field of study of this research-action is composed of a catalogue of 800 non-fiction movies (offered to a film-loving public by the video-on-demand platform Tënk) and the exploitation of the Rumo algorithmic engine (made available by the company Spideo). In a first step, a corpus of 112 films was indexed via a thesaurus of filmic processes produced specifically for this study. In a second step, the results of the algorithmic recommendation were compared to a recommendation established by a panel of experts specialized in the field. The results highlight new networks of relevant links between films that did not previously exist with the thematic classification. They allow us to sketch out new perspectives for improving the digital mediation of documentary films on platforms.

Keywords. Documentary cinema, Video-On-Demand platform, audiovisual indexing, recommendation algorithm, proof of concept, User eXperience Design, digital mediation.

Introduction

Depuis une décennie, l'offre légale de vidéo à la demande s'est développée massivement sur Internet. Aujourd'hui, quiconque souhaitant visionner un film documentaire fait face à un double défi¹ : premièrement, adopter une plateforme parmi une grande variété d'offres en ligne² et, deuxièmement, arrêter son choix de visionnage en fouillant des catalogues de plusieurs milliers de titres. Face à cet hyperchoix, les acteurs économiques des industries culturelles développent des dispositifs d'aide à la décision (Labour & Abou Jaoudé, 2014) pour renforcer l'engagement de leur audience. Ces dispositifs de recommandation sont de trois ordres : éditorial (curation des œuvres par des programmeurs), social (interaction entre usagers des plateformes) et algorithmique (filtrage automatisé des catalogues).

Dans ce contexte, cet article présente les résultats d'une expérimentation menée en 2020 dans le cadre du programme AlgoDoc (Algorithme de recommandation de films documentaires)³. Son objectif est d'expérimenter une recommandation basée sur une nouvelle forme de catégorisation du cinéma documentaire alternative aux métadonnées fondées principalement sur une classification thématique. Pour ce faire, un thésaurus qualifiant la grande variété des dispositifs de réalisation a été créé afin d'indexer un corpus de 112 films représentatif du catalogue de la plateforme Tënk⁴. Dans un deuxième temps, le

¹ Conscient de cette problématique, le CNC a développé un moteur de recherche permettant de filtrer les catalogues de l'offre légale de vidéo à la demande : <http://vad.cnc.fr/> (consulté le 02/12/2020).

² Selon le rapport établi en 2021 par Hadopi et le CSA, au 30 avril 2020, 78 services de vidéo à la demande sont recensés en France.

³ Ce programme a été financé en 2020 par l'I-Site et la Satt Nord. Il réunit un consortium recherche-industrie composé des laboratoires GERiCO de l'Université de Lille et DeVisu de l'Université Polytechnique Hauts-de-France, ainsi que la plateforme de films documentaires Tënk et la société Spideo, spécialisée dans la recommandation algorithmique de contenu audiovisuel.

⁴ <https://www.tenk.fr> (consulté le 07/12/2020).

travail a consisté à évaluer la pertinence d'une recommandation automatisée via un moteur algorithmique mis à disposition par la société Spideo⁵ puis à comparer les résultats avec un panel d'experts du domaine. Les questions de recherche soulevées par cette étude sont multiples : quelles métadonnées faut-il mobiliser pour qualifier les films indépendamment des thématiques abordées et de leurs auteurs ? Les dispositifs de réalisation⁶ utilisés par les cinéastes permettent-ils d'indexer de manière pertinente le catalogue de Tënk ? À quelle condition la computation algorithmique de ces métadonnées permet-elle d'établir des réseaux de liens pertinents entre les films ?

Notre étude s'inscrit dans la continuité d'une « recherche-projet en design » (Findeli, 2013 [2006], 2015 ; Gentès, 2015, 2017 ; Masure, 2017)⁷ en adoptant une méthodologie de conception exploratoire, itérative et réflexive focalisée sur la résolution de problème (Gaudenzi & Gantier, 2018 ; Zacklad, 2017). Le processus de recherche vise à « faire pour comprendre » davantage que de « comprendre pour faire ». L'approche interdisciplinaire adoptée s'inscrit à la croisée du design d'expérience utilisateur ; des études cinématographiques ; des sciences de l'information et de la documentation ; et des sciences informatiques. Comme l'argumente Annie Gentès, l'intérêt d'une recherche interdisciplinaire par le design est de renouveler les apports de connaissance sur les innovations d'usage dans les TIC – chaque discipline amenant avec elle sa propre représentation de l'utilisateur. Cette approche se veut « avant tout générique et non pas cumulative. Il ne s'agit pas seulement d'emprunts à différentes théories et méthodes, il s'agit de construire un objet nouveau, un X, et cet X permet de reformuler les théories et les méthodes » (Gentès, 2015 : 105).

La première partie de l'article présente les différents paradigmes sur lesquels s'appuie cette recherche par le design. Afin de mettre au jour les postulats préexistants à la conception de la preuve de concept, il nous a semblé primordial d'explicitier dans un premier temps les besoins de l'équipe éditoriale de Tënk ; les attentes des utilisateurs cinéphiles de la plateforme ; ainsi que le positionnement stratégique et commercial de la société Spideo. La deuxième partie détaille la méthodologie adoptée pour la conception et l'évaluation de la preuve de concept (critères retenus pour construire le corpus ; processus d'indexation et d'ingestion des données ; recueil des recommandations auprès d'un panel d'experts du domaine). La troisième partie analyse les écarts entre la recommandation algorithmique et la prescription humaine des experts. Enfin, les apports et limites de cette preuve de concept sont discutés afin d'identifier des pistes de recherches ultérieures susceptibles d'améliorer la médiation algorithmique du cinéma documentaire sur les plateformes.

⁵ <https://spideo.com/en/> (consulté le 07/12/2020).

⁶ Par souci de commodité, nous qualifierons dans cet article l'ensemble des enjeux esthétiques et communicationnels propres au cinéma documentaire de « dispositifs de réalisation ». Cette notion est employée de manière synonyme avec l'idée de procédés filmiques ou de parti-pris de réalisation.

⁷ Sur la manière dans la recherche en design renouvelle les concepts cardinaux en SIC de médiation, médiatisation et usages des TIC, voir notamment l'ouvrage collectif de la CPDirSIC (2019, p. 143-54).

1 Enjeux épistémologiques d'une recherche par le design

1.1 Paradigmes mobilisés par l'équipe de recherche

Cette partie ne prétend pas dresser un état de l'art exhaustif des très nombreuses études menées, depuis une dizaine d'années, sur les algorithmes de recommandation mais vise davantage à préciser les différents paradigmes sur lesquels s'appuie cette étude. À cette fin, trois grandes familles de travaux ont été sollicitées.

Premièrement, des études en science de l'information et de la communication questionnent les modes de prescription techno-sémiotique mis en œuvre par les acteurs des industries culturelles pour valoriser leurs catalogues (Beuscart *et al.*, 2019 ; Bullich, 2018 ; Chapelain & Ducas, 2018 ; Croissant, 2019 ; Croissant & Cambone, 2020 ; Debryne & Pirolli, 2017 ; Drumont *et al.*, 2018 ; Farchy *et al.*, 2017 ; Paquiénéguy, 2017 ; Riondel, 2020 ; Taillibert, 2020). Ces recherches présentent l'intérêt de replacer les plateformes de vidéo à la demande dans un ensemble plus vaste de médiations instrumentées de biens culturels. Ils positionnent le design d'expérience utilisateur de la plateforme au cœur d'un système de recommandation qui articule trois logiques complémentaires : une prescription éditoriale (des programmeurs vers les abonnés), une prescription sociale (des abonnés vers les abonnés) et une prescription algorithmique centrée sur le contenu (des films vers les films).

Deuxièmement, des recherches en sociologie des techniques interrogent de manière critique les dispositifs de médiation informatisée. Ils mettent en exergue leur rôle dans la mise en données par les acteurs économiques des différentes sphères de la société (Amer-Yahia *et al.*, 2020 ; Cardon, 2013, 2015 ; Ertzscheid, 2015 ; Gillespie, 2014 ; Ménard, 2014 ; Musiani, 2013 ; Vayre, 2017). Dans un contexte de collecte massive des données personnelles, ces travaux déconstruisent la pseudo-neutralité des algorithmes, leur usage par les plateformes d'intermédiation, et soulignent leurs nombreux biais et impacts sociaux. Comme le souligne Dominique Cardon : « les calculateurs fabriquent notre réel, l'organisent et l'orientent. Ils produisent des conventions et des systèmes d'équivalence qui sélectionnent certains objets au détriment d'autres, imposent une hiérarchisation des valeurs qui en vient progressivement à dessiner les cadres cognitifs et culturels de nos sociétés » (Cardon, 2015 : 14). Suivant le même raisonnement, Olivier Ertzscheid (2015) appelle à un travail de conscientisation par l'explicitation des calculs et données associées (phénomènes de rétention, masquage, et enfouissement qui s'opposent au fait de divulguer, révéler, dévoiler ce processus computationnel). En filiation avec ces travaux, cette étude a pour ambition d'ouvrir la « boîte noire » du dispositif socio-technique de recommandation afin de disséquer son mécanisme interne et de mieux appréhender sa pertinence pour prescrire des films documentaires aux abonnés de Tënk. En effet, si les impacts sociaux des algorithmes ne sont plus à démontrer, le coût pour concevoir une alternative aux systèmes algorithmiques développés par les sociétés dominantes du marché reste un défi de taille pour de petits acteurs des industries culturelles tels que la plateforme Tënk.

Troisièmement, des recherches à la jonction des sciences de l'information et de la documentation et des sciences informatiques focalisent leur analyse sur l'évaluation des modèles mathématiques et leurs impacts sur les usages (Bertail *et al.*, 2019 ; Domenget & Coutant, 2014 ; Foulonneau *et al.*, 2014 ; Hachour *et al.*, 2014 ; Linden *et al.*, 2003 ; Ricci *et al.*, 2010 ; Rieder, 2020 ; Schmitt, 2016). Olessia Kirtchik

en propose une lecture critique intéressante (Kirtchik, 2019). De façon générale, ces recherches soulignent qu'il est tout aussi important d'interroger les biais dans la construction des données que le processus computationnel en lui-même. Elles définissent les systèmes automatisés de recommandation selon trois grands paradigmes (Dahimene, 2014 ; Kembellec *et al.*, 2014) :

- Les approches centrées sur les données des utilisateurs (filtrage collaboratif) se basent sur l'hypothèse que les personnes qui ont aimé des contenus identiques par le passé possèdent un goût similaire et sont propices à apprécier les mêmes contenus dans le futur. Basées sur l'ensemble des interactions des utilisateurs, ces méthodes présentent l'avantage de recommander des objets complexes, sans avoir à les analyser en amont⁸. Parmi les nombreuses contributions de ce type d'approches (Ahn & Shi, 2009 ; Bernier *et al.*, 2011 ; Kouadria & Nouali, 2014 ; Lherisson, 2018), Shinhyun Ahn et Chung-Kon Shi utilisent notamment les métadonnées culturelles telles que les critiques, les évaluations et les commentaires sur les blogs et les sites de médias sociaux pour créer leur système (Ahn & Shi, 2009).
- Les démarches centrées sur le contenu exploitent des métadonnées qualifiant de larges catalogues qu'elles mettent en correspondance avec un profilage de l'utilisateur. Ces approches nécessitent un travail préalable d'indexation sémantique des contenus et impliquent d'adopter des facettes discriminantes. Parmi les travaux existants, Sejwal et Abulaish (2021) proposent une approche basée sur une ontologie de domaine caractérisant le contenu de films (genre, mots clés, etc.). Les auteurs construisent l'ontologie à partir de différentes bases de données de films disponibles sur le web. Néanmoins, cette ontologie ne peut être utilisée dans notre expérimentation car elle n'est pas définie pour caractériser des films documentaires tels que ceux disponibles sur la plateforme Tënk.
- Les approches hybrides combinent les démarches centrées sur l'utilisateur et le contenu⁹. Parmi les travaux existants, Walek et Fojtik (2020) proposent un système de recommandation monolithique appelé Predictory. Celui-ci travaille avec les genres favoris et impopulaires de l'utilisateur, tandis que la liste finale des films recommandés est déterminée à l'aide d'un système « expert flou », qui évalue l'importance des films. Le système travaille avec plusieurs paramètres provenant de bases de données du web : l'évaluation moyenne des films, le nombre d'évaluations, et le niveau de similarité entre les films déjà évalués. Bien qu'intéressante, cette approche n'est pas adaptée au catalogue de documentaires de création de Tënk pour lequel il n'existe pas ces jeux de données accessibles en ligne. Pour aller plus loin, les travaux de Roza Lemdani ou encore ceux de Robin Burke dressent un état détaillé des approches hybrides pour la recommandation (Burke, 2007 ; Lemdani, 2016).

Dans tous les cas, quel que soit le type de méthodologie adoptée, l'algorithme de recommandation nécessite un volume important de données issues d'interactions des utilisateurs ou de la description des contenus pour pouvoir amorcer le système

⁸ L'un des exemples canoniques est l'algorithme *Item-to-item Collaborative Filtering* développé en 1998 par Amazon (Linden *et al.*, 2003).

⁹ C'est le cas notamment du célèbre algorithme de Netflix qui associe l'analyse des facettes (genre et sous genre) avec les traces d'usage des abonnés (Amatriain, 2013 ; Drumont *et al.*, 2018).

de prescription – base d'apprentissage nécessaire pour lever le problème bien connu du « démarrage à froid ».

1.2 Besoins de l'équipe éditoriale de Tënk

Créée en 2016, la plateforme Tënk revendique, en décembre 2020, plus de 20 000 abonnés (Gantier, 2020). Dans un marché fortement concurrentiel et structuré en oligopole à franges (Thuillas & Wiart, 2019), la plateforme cherche à se différencier à travers une approche artistique forte en opposition aux programmes télévisés standardisés (Lesauvier, 2019). Sur le plan éditorial, la plateforme propose une « programmation glissante » (Taillibert, 2020) d'une cinquantaine de films documentaires accessibles pendant 2 mois sur abonnement (SVOD) auxquels s'ajoute un catalogue de près de 800 titres visionnables à l'acte (VOD). En 2020, cette programmation est structurée en différentes rubriques thématiques intitulées « plages » (histoire et politique, arts, écologie, musique, sciences, festivals, etc.). L'ensemble des films sont sélectionnés par une équipe de programmeurs bénévoles qui revendiquent de les choisir pour « la subjectivité d'un regard¹⁰ » s'incarnant dans un point de vue singulier sur le monde et une démarche artistique assumée. En faisant de la défense de la création documentaire contemporaine et patrimoniale le socle de son identité, la plateforme cible prioritairement un public cinéphile appréciant particulièrement le « cinéma du réel ». Pour le directeur de la plateforme, l'objectif « est de mieux accompagner le public initié mais également non-initié au documentaire d'auteur. L'enjeu est de faire évoluer les représentations négatives sur le genre documentaire qui est trop souvent associé au film animalier, pédagogique ou au reportage journalistique¹¹ ». Dans ce but, l'équipe travaille sur une refonte de la plateforme dont le lancement est prévu en 2021 :

« Nous réfléchissons à la manière de pouvoir dépasser une première expérience de visionnage décevante si le spectateur n'a pas aimé le premier film qu'il a vu et qu'il se dit : le documentaire d'auteur ce n'est pas pour moi ! (...). La création d'une recommandation algorithmique en complément de notre éditorialisation très forte et de la recommandation des abonnés doit nous permettre, à terme, de renforcer le renouvellement des abonnements. Un deuxième chantier consiste à refondre notre offre commerciale de manière à mieux valoriser notre offre VOD. Nous avons acquis les droits d'environ 800 films depuis 4 ans mais nous n'avons pas eu le temps ni les moyens pour travailler sur la monétisation du catalogue¹² ».

Ce constat conduit l'équipe éditoriale à vouloir imaginer un jeu de données traduisant la grande variété des dispositifs de réalisation qui composent l'espace du cinéma documentaire. De manière générale, les plateformes généralistes indexent leur catalogue en qualifiant différentes facettes de métadonnées : auteur-réalisateur, comédiens, date, thématique, genre cinématographique (Delaporte, 2019). Or, ces taxinomies s'avèrent inopérantes lorsque l'on focalise sur le segment restreint du cinéma documentaire d'auteur. En effet, ces catégories ne permettent pas de rendre compte de la diversité des démarches cinématographiques qui composent l'espace du cinéma documentaire (Gauthier, 2015 [1995] ; Kilborne, 2008 ; Lioult, 2019 ; Nichols, 2010 [2001] ; Ninoy, 2009). C'est pourquoi le processus de conception a

¹⁰ Expression extraite de l'éditorial du site : <https://www.tenk.fr/p/notre-projet> (consulté le 27/01/2020).

¹¹ Extrait d'entretien mené avec Pierre Mathéus, directeur général de Tënk, recueilli à Lussas le 06 juillet 2020.

¹² *Ibid.*

débuté par la création d'un thésaurus inédit permettant de caractériser finement les dispositifs filmiques (Givois *et al.*, 2021). L'hypothèse centrale de ce jeu de métadonnées est qu'il permettrait de dépasser la simple documentation thématique afin de faire émerger de nouveaux réseaux de liens entre les films du catalogue.

1.3 Attentes et frustrations des abonnés cinéphiles de Tënk

Une étude exploratoire, menée en amont du design de la preuve de concept, a permis de récolter de nombreuses données quantitatives et qualitatives sur les utilisateurs cinéphiles de Tënk (Gantier, 2020). La construction de *personae*¹³ vise à qualifier, de manière compréhensive, les motivations, frustrations, obstacles, besoins et attentes de différents groupes d'abonnés de la plateforme. Deux éléments ressortent de cette formalisation des usages. Tout d'abord, le premier cercle d'abonnés, constitué principalement de cinéphiles et de praticiens du cinéma documentaire, se focalise moins sur la thématique abordée par les films (leur sujet), que sur son dispositif de réalisation (leur traitement cinématographique). D'autre part, l'offre de 50 films disponibles pendant 2 mois représente une quantité de propositions difficile à assimiler. Cette surabondance de choix est renforcée par un catalogue d'environ 800 films en visionnage à l'acte. Plusieurs *personae* témoignent de leurs frustrations face à cette offre artistique pléthorique. C'est le cas par exemple de Léonard L. qui déclare : « Chaque semaine vous proposez des films riches mais c'est presque trop et il est parfois difficile de se souvenir de ce qu'on a vu. Même un retraité à plein temps, ce qui n'est pas mon cas, n'arriverait pas à tout voir dans le temps imparti ! ». Julien P. regrette quant à lui un manque d'accompagnement pour l'aider dans sa sélection : « Je suis un peu perdu face à l'offre abondante. Je n'ai pas les références et ne sais pas toujours quoi regarder, du coup je zappe pas mal... ». La proposition éditoriale de la plateforme s'inscrit ainsi dans un environnement fortement concurrentiel où les différents acteurs des industries audiovisuelles cherchent à capter et conserver l'attention de leur audience sur Internet (Citton, 2014).

Dans ce contexte, la création d'un jeu de métadonnées permettant de qualifier sémantiquement les dispositifs filmiques du catalogue de Tënk tente de répondre à un double défi. Il s'agit, d'une part, d'établir une correspondance entre la nature des données importées dans l'algorithme et les spécificités cinématographiques du catalogue, ceci afin que la médiation algorithmique s'inscrive en parfaite cohérence éthique, artistique et politique avec la promesse éditoriale de la plateforme. L'objectif est, d'autre part, de mieux accompagner les abonnés cinéphiles dans leur choix de visionnage. Cette aide à la décision doit « faire sens » pour un public cinéphile qui connaît et apprécie la grande diversité des démarches cinématographiques qui composent le champ du cinéma documentaire.

1.4 Positionnement stratégique et commercial de Spideo

La société Spideo, composée d'une équipe d'une vingtaine de salariés, fournit des dispositifs de recommandation à des télévisions et plateformes VOD¹⁴. Son

¹³ Les *personae* sont « des archétypes d'utilisateurs créés à partir de données réelles recueillies pendant la phase d'exploration. [Et] sont utiles pour générer des idées et guider dans des décisions de conception » (Lallemand & Gronier, 2018 [2015] : 194).

¹⁴ La société travaille pour deux types de clients : des opérateurs télécoms qui commercialisent des catalogues audiovisuels (Bouygues, SFR) et des médias historiques qui proposent des offres délinéarisées (Canal +, France Télévisions, Televisa, Sky, etc.). Son objectif, à moyen terme, est de proposer également une offre de services moins onéreuse pour les nouveaux acteurs OTT qui valorisent une offre culturelle de niche.

positionnement stratégique et commercial sur le marché des dispositifs d'intermédiation consiste à revendiquer une approche sémantique pour indexer de larges catalogues audiovisuels multi-genres (fiction, série, animation, œuvres patrimoniales et programmes de flux)¹⁵. Son co-fondateur et gérant, Thibault d'Orso, définit son métier comme la version numérique « du vendeur-conseiller d'un vidéoclub [dont la] fonction est de faire découvrir des œuvres dans un dispositif conversationnel¹⁶ ». Dans la mesure où les industries culturelles proposent de faire « des expériences singulières », il ne lui semble pas pertinent de dupliquer les mécaniques marketing déployées massivement dans le secteur du e-commerce.

« Dans le secteur culturel, il est difficile de se contenter d'une phrase du type : d'autres consommateurs comme vous ont aussi acheté... Ce n'est pas seulement que c'est frustrant ou décevant mais c'est insultant dans un secteur aussi porteur de signaux et d'enjeux d'identification. (...) Autrement dit, la particularité de ce qu'on fait : la sémantique, c'est ce qui permet de ré-humaniser tout un pan très technique des systèmes de recommandation qui, pendant tout un temps, appartenait essentiellement au domaine de la statistique¹⁷ ».

Cette approche sémantique, centrée sur le film, s'oppose à un traitement statistique, centré sur les traces de visionnage des utilisateurs. En effet, « l'algorithmique statistique ne fait que recycler une information purement endogène à la sphère de l'utilisateur, [alors que] l'algorithmique sémantique apporte une expertise exogène, incorporée dans l'indexation préalable des contenus » (CSA Lab, 2017, p. 17). Toutefois, il faut souligner que le gain qualitatif dans le traitement de l'information présente une contrepartie importante, en termes de coût de développement, dans la mesure où les métadonnées sémantiques doivent être générées *ex-nihilo* par l'instance d'énonciation elle-même. Sa mise en œuvre est donc plus onéreuse car elle exige en amont de la mise en calcul un travail d'expertise et d'indexation conséquent, contrairement aux algorithmes statistiques déployés seuls qui exploitent des données préexistantes à travers les traces laissées par les usagers des plateformes et réseaux sociaux.

Le discours d'escorte du moteur algorithmique commercialisé par Spideo (tel qu'on peut le trouver sur son site Internet) met en avant une recommandation qui remplace l'expertise humaine au cœur d'un dispositif communicationnel en langage naturel :

« La technologie Spideo relie chaque élément de contenu à une empreinte sémantique unique basée sur des groupes de mots-clés pondérés et catégorisés, ce qui permet une découverte au-delà des genres de base. (...) Les mots clés sont basés sur le langage naturel pour classer le contenu de nos clients. Ils alimentent nos algorithmes de recommandation pour aider les spectateurs et les cinéphiles à découvrir, non seulement

¹⁵ Sur le marché de l'ingénierie de la recommandation, Spideo est en concurrence avec des sociétés comme Aws (Amazon web services), ViaAccessOrca (filiale du groupe Orange), Dynamic Yield, Yusp ou CogniK qui proposent leurs moteurs de recommandation aux acteurs du e-commerce comme aux industries culturelles.

¹⁶ Entretien réalisé par la Turbine.coop à Grenoble en septembre 2019 dans le cadre d'une journée sur les algorithmes des industries culturelles : <https://turbine.coop/2019/09/la-turbine-coop-lance-ses-podcasts-retour-sur-culture-futur-2019-1-2-une-exploration-des-algorithmes-de-recommandation-culturelle> (consulté le 08/12/2020).

¹⁷ *Ibid.*

des contenus similaires, mais aussi des listes thématiques de contenus connexes triés par catégories pertinentes et personnalisées¹⁸ ».

Au regard de l'état de l'art sur les dispositifs de recommandation présenté précédemment (cf. partie 2.1) cette approche sémantique n'est pas nouvelle mais s'inscrit plutôt dans une longue tradition historique des techniques d'ordonnement de l'information (Rieder, 2020). En effet, le moteur intègre une méthode par filtrage de contenu qui s'appuie sur un prétraitement préalable manuel et vise à décrire les contenus audiovisuels par différentes facettes. Il s'agit d'une «procédure de classification supervisée [qui] ne produit pas de résultats mécaniquement sans intervention humaine. (...) L'ingénieur intervient sur les données, les procédures, et les ajustements possibles en essayant d'améliorer les résultats par tâtonnements» (Schmitt, 2016 : 93). L'apport de sens préconisé par Spideo se caractérise ainsi par un travail d'expertise artisanal important qui cherche à améliorer la recommandation en amont et en aval de la computation des données.

2 Processus de conception d'une recommandation centrée sur les dispositifs de réalisation des films documentaires

2.1 Design d'une preuve de concept

L'expérimentation menée dans cette étude répond à un cheminement itératif que l'on peut décomposer en sept grandes étapes, comme le détaille la figure ci-dessous.

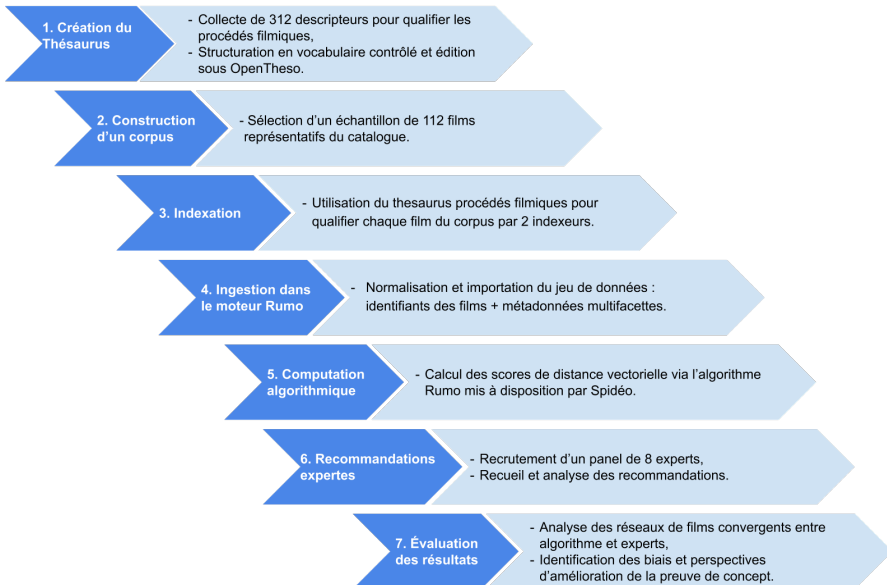


Figure 1. Processus de conception de la preuve de concept

¹⁸ <https://spideo.com/en/semantic-data/> (consulté le 14/12/2020).

2.2 Construction d'un corpus représentatif du catalogue de Tënk

Parallèlement à la création d'un thésaurus de dispositifs de réalisation du cinéma documentaire, un corpus de films a été constitué afin d'indexer une partie représentative du catalogue de Tënk. À cette fin, 112 films ont été sélectionnés parmi les 1 414 documentaires dont la plateforme a acquis les droits de diffusion en avril 2020. Ce corpus représente 8% du catalogue et se compose de courts et longs métrages, d'œuvres contemporaines et patrimoniales, francophones et internationales. Une première sélection aléatoire ayant entraîné une surreprésentation de films courts et récents, un redressement de l'échantillon a ensuite été effectué afin d'améliorer sa représentativité. Celui-ci a été mené en fonction de deux critères factuels : la date de production (figure 2) et la durée du film (figure 3).

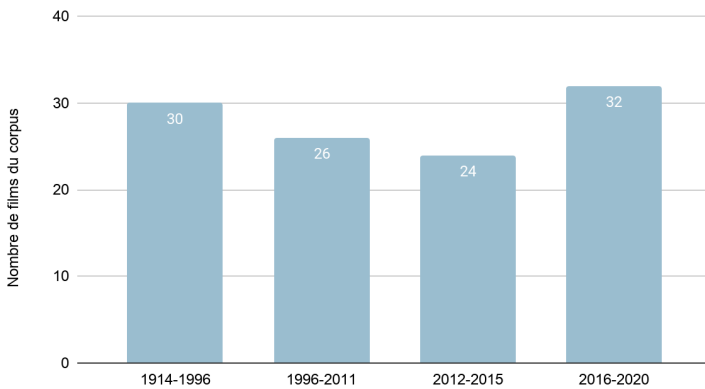


Figure 2. Distribution du corpus selon les quartiles de date de production du catalogue de Tënk

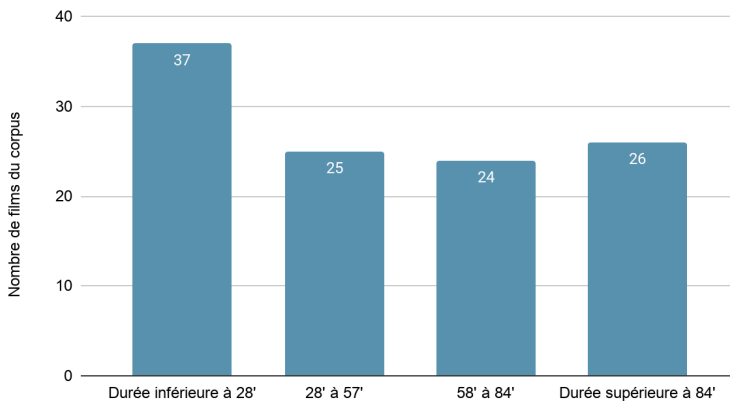


Figure 3. Distribution du corpus selon les quartiles de durée du catalogue de Tënk

S'appuyant sur l'hypothèse que les dispositifs filmiques sont fortement situés historiquement, la construction du corpus s'est attachée à représenter la distribution du catalogue en termes de période de réalisation (date) et de format de production (durée). En effet, selon les périodes considérées, les formats de diffusion cinématographique et télévisée, les structures de production ainsi que les outils et dispositifs de réalisation socio-techniques évoluent fortement (avènement des caméras argentiques dotées d'un dispositif d'enregistrement du son synchrone, miniaturisation de la vidéo numérique ou développement des caméras connectées, par exemple).

2.3 Indexation itérative du corpus

Une indexation a ensuite été effectuée à l'aide du thésaurus des dispositifs de réalisation du cinéma documentaire, créé spécifiquement pour cet usage¹⁹. Ce thésaurus, composé de 5 branches et 312 descripteurs, a été structuré en adoptant une terminologie conceptuelle. Cette approche a été choisie car elle permet de rendre compte des paradigmes communicationnels mobilisés dans les typologies des modes de représentation du réel proposées par Nichols (2010 [2001]) et Ninoy (2009). L'objectif est de parvenir à expérimenter une nouvelle forme de catégorisation du cinéma documentaire alternative aux métadonnées traditionnelles basées principalement sur une classification thématique. Ainsi, afin de favoriser des rapprochements sémantiques pertinents au sein du corpus, chaque film a été indexé de manière à qualifier finement les dispositifs les plus significatifs, c'est-à-dire les éléments formels et communicationnels qui rendent compte des partis-pris de réalisation ou dispositifs filmiques de leurs auteurs²⁰.

Deux limites ont été identifiées dans l'utilisation de ce thésaurus. Premièrement, la manipulation des descripteurs conceptuels n'est pas aisée et peut conduire à des biais liés à la subjectivité des indexeurs. Pour limiter cette lecture interprétative du thésaurus, le codage a été réalisé parallèlement par deux indexeurs autonomes, chaque film bénéficiant ainsi d'un double regard pour valider la liste de termes retenue. Deuxièmement, le niveau de langue spécialisé ne permet pas à l'abonné de manipuler le vocabulaire contrôlé directement lors de ses recherches de films sur la plateforme. Un travail de traduction et de simplification de cet outil est donc à prévoir dans des recherches ultérieures afin de pouvoir imaginer un usage industriel.

2.4 Ingestion du jeu de données

L'étape suivante a consisté à importer un jeu de données dans le moteur algorithmique de Spideo, accessible via leur outil Rumo. Celui-ci est constitué des identifiants des films et des métadonnées associées, préalablement normalisées au modèle EBU Core²¹. Ces métadonnées se décomposent en 36 facettes (titre, auteur, résumé, avis, année de production, pays de production majoritaire, public familial, etc.). Dans le cadre de cette preuve de concept, seule la facette des dispositifs de

¹⁹ La construction méthodologique de ce thésaurus est détaillée dans une publication antérieure (Givois *et al.*, 2021). Cet outil est accessible, en données ouvertes, sous OpenTheso (<https://opentheso.huma-num.fr/opentheso/?idt=th228>).

²⁰ Le rapport scientifique et technique du programme AlgoDoc consigne un guide d'indexation, ainsi que la liste exhaustive des 112 films du corpus avec leurs métadonnées thématiques et dispositifs filmiques associées (Gantier *et al.*, 2020).

²¹ Ce modèle a été choisi pour son interopérabilité (données ouvertes, disponible aux formats XML, JSON et SKOS) : <https://tech.ebu.ch/MetadataEbuCore>, consulté le 10/12/2020.

réalisation a été exploitée afin de pouvoir analyser ses effets sur la recommandation algorithmique.

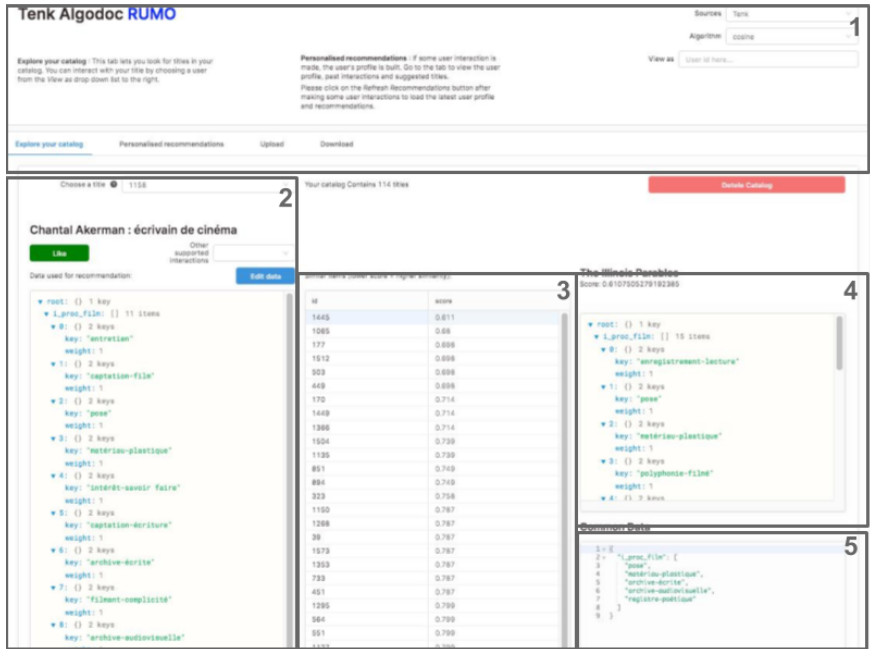


Figure 4. Capture d'écran du tableau de bord de Rumo

La figure ci-dessus présente le tableau de bord du moteur algorithmique Rumo²² qui se décompose en cinq zones distinctes. La première partie permet de charger une collection de films et de sélectionner un algorithme de comparaison des documents parmi une bibliothèque de ressources en open access. Le reste de l'interface présente les métadonnées associées à chaque film (zone 2) et permet de comparer sa proximité avec les autres films du corpus (zones 3 à 5). Plus précisément, la fenêtre 3 affiche les scores des films dont la distance vectorielle avec le documentaire sélectionné est la moins élevée. La zone 4 détaille quant à elle les descripteurs du film sélectionné en zone 3. Enfin, la fenêtre 5 montre les métadonnées partagées par les films sélectionnés en zone 2 et 3. Cette interface de contrôle permet ainsi à l'équipe de recherche de visualiser les descripteurs associés à un film et de mieux appréhender la manière dont leur pondération est prise en considération par le calcul algorithmique. Cet outil intermédiaire entre l'équipe de recherche et le corpus constitue un véritable architecte, à savoir un méta-texte « qui, non content de représenter la structure du texte, en commande l'exécution et la réalisation » (Jeanneret & Souchier, 1999 : 105-8). Ce concept désigne des objets informatiques qui sont en position de régir l'écrit, c'est-à-dire de lui attribuer ses

²² Moteur en phase de développement par Spideo pendant l'étude et testé via l'application Retool.

formats et ses ressources. Il s'agit, en quelque sorte, d'une « écriture de l'écriture, d'une écriture qui conditionne d'autres écritures » (Tardy & Jeanneret, 2007 : 24).

2.5 Computation algorithmique

Les algorithmes de comparaison des documents proposés par Rumo exploitent la technique de traitement automatique VSM (Vector Space Model) très répandue dans les technologies d'extraction d'information²³. Chaque document est défini par un vecteur signature qui *spatialise* un ensemble de termes descripteurs. En ce qui concerne cette preuve de concept, le vecteur représente la facette de métadonnées construite lors de l'indexation des dispositifs de réalisation. Chaque vecteur signature d'un document est ensuite comparé à l'ensemble des vecteurs du corpus par l'intermédiaire des mesures disponibles dans la bibliothèque. La métrique Cosinus²⁴ (cf. figure 6) a été préférée à celle de l'indice de Jacquard car elle présente deux avantages majeurs. Premièrement, elle prend en compte la pondération des termes constituant un vecteur – pondération nécessaire pour exprimer qu'un dispositif est plus important qu'un autre pour exprimer une démarche cinématographique particulière. Deuxièmement, elle pénalise moins le résultat lorsque les vecteurs comparés sont de tailles très différentes, c'est-à-dire lorsque l'on compare un film indexé avec peu de termes avec un autre qualifié avec beaucoup de descripteurs. Néanmoins, cette mesure reste relativement inefficace lorsque les vecteurs comparés sont de trop grande taille (car cela augmente mécaniquement le nombre de termes distincts).

$$\text{sim}(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

Figure 5. Définition du Cosinus de Salton

Dans un premier temps, l'expérimentation a consisté à tester l'algorithme avec des vecteurs signatures définis sans aucune restriction de taille. Lors de la phase d'indexation, chaque descripteur caractérisant, de près ou de loin, un dispositif cinématographique, permet d'enrichir le vecteur signature. Ensuite, chaque vecteur a été comparé aux autres selon la mesure cosinus. Appliquée à notre corpus de 112 films documentaires, cette méthodologie donne des résultats mitigés. Ceci s'explique notamment par le fait que les vecteurs signatures sont de tailles très différentes (allant de 3 à plus de 20 descripteurs par films). En effet, les vecteurs de

²³ Bernhard Rieder, dans son ouvrage *Engine of Order: a Mechanology of Algorithmic Techniques*, dresse une archéologie des techniques d'ordonnement algorithmique. Il détaille notamment la genèse du concept de vecteur caractéristique ou vecteur signature qui constitue une structure de données canonique pour de nombreuses techniques d'extraction de l'information depuis son invention par Gerard Salton dans les années 1960 (Rieder, 2020 : 199-233).

²⁴ Le calcul de la métrique Cosinus (dit Cosinus de Salton) est basé sur les mesures de la similarité du vecteur par la formule des vecteurs cosinus. Soit deux vecteurs A et B, l'angle θ s'obtient par le produit scalaire et la norme des vecteurs. On cherche ici à obtenir la valeur $\cos \theta$ qui est comprise dans l'intervalle [-1,1]. La valeur -1 indiquera des vecteurs résolument opposés, la valeur 0 des vecteurs indépendants (orthogonaux) et la valeur 1 des vecteurs similaires (colinéaires de coefficient positif). Les valeurs intermédiaires permettent d'évaluer le degré de similarité.

grandes tailles favorisent un plus grand nombre de descripteurs différents et ajoutent par là-même du bruit dans le calcul de proximité entre les films. Pour pallier ce problème, nous avons cherché à harmoniser le nombre de termes associés à chaque film, ainsi qu'à identifier, par tâtonnement successif, la taille vectorielle la plus efficace pour comparer des films les uns aux autres. L'enjeu a donc consisté à hiérarchiser les descripteurs des dispositifs les plus emblématiques de la démarche cinématographique d'un film. Pour ce faire, un ordonnancement a été établi pour limiter progressivement le nombre de termes de 20 à 8, 5 et 3 descripteurs. Le parti pris méthodologique retenu a été de normaliser un jeu de données constitué de 8 descripteurs par film pour l'ensemble du corpus indexé.

3 Construction d'une recommandation humaine du cinéma documentaire

3.1 Recueil d'une recommandation experte

Afin de pouvoir évaluer la pertinence de la recommandation algorithmique, la deuxième partie de l'étude a consisté à élaborer une recommandation humaine. À cette fin, un panel composé de 8 experts du cinéma documentaire a été recruté²⁵. Ce groupe de spécialistes cinéphiles possède une culture de l'image relativement homogène dans la mesure où tous ont fréquenté régulièrement les grands festivals du secteur depuis une vingtaine d'années. De plus, la totalité du panel a déjà réalisé un documentaire de création, possède une expérience de programmation et apprécie la grande diversité des dispositifs qui composent le champ hétérogène du cinéma documentaire.

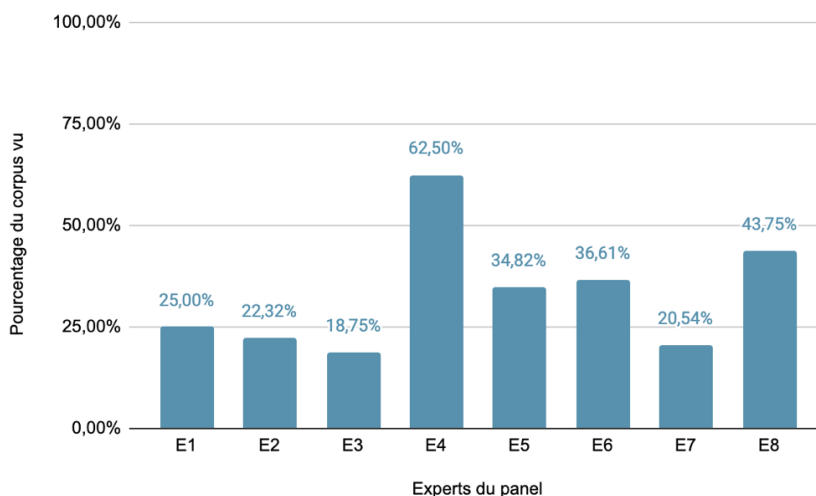


Figure 6. Distribution du panel en fonction des films du corpus vus

²⁵ Le panel se compose de 3 membres de l'équipe éditoriale de Tënk, 3 enseignants-chercheurs en études cinématographiques et 2 réalisateurs-formateurs.

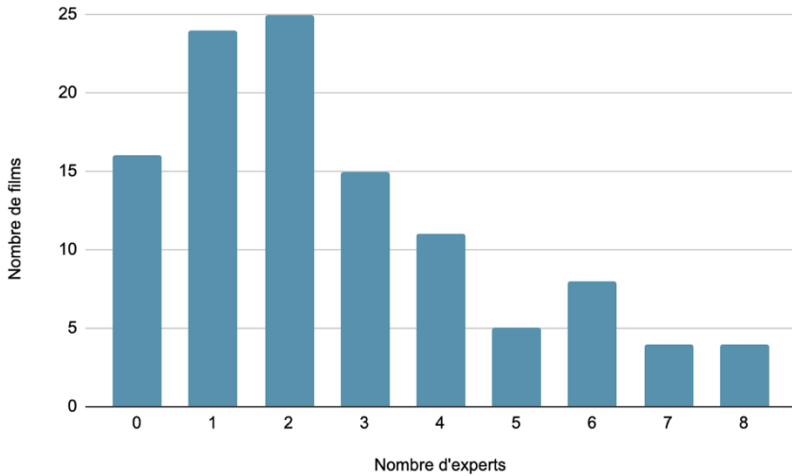


Figure 7. *Quantité de films vus par n experts*

Le recueil de la recommandation des experts représente un défi méthodologique dans la mesure où aucun humain, expert du domaine comme programmeur de T'enk, ne peut visionner, en amont de l'expérimentation, l'intégralité des 112 films du corpus (qui représente plusieurs centaines d'heures de visionnage). De fait, sur les 112 films du corpus, 16 films n'ont pas été visionnés par le panel et 72 films ont été vus par au moins 2 experts. Dès lors, les résultats de cette recommandation humaine constituent un construit scientifique « par essence foncièrement subjectif et asservi à un projet donné » (Leleu-Merviel, 2008 : 20). Cette prudence épistémologique nous a conduit à effectuer un recueil de données en suivant trois étapes successives. Dans un premier temps, un formulaire a permis à chacun des experts de déclarer les films qu'ils avaient vus et dont ils se souvenaient. Dans un deuxième temps, le panel a disposé d'une semaine pour renseigner une grille de correspondance parmi les films visionnés. La consigne était d'identifier uniquement les films présentant des similarités de dispositifs en excluant tous les autres critères de rapprochement (thématique, auteur, période, etc.). Enfin, un entretien d'auto-confrontation a été mené pour expliciter l'ensemble des liens déclarés par les experts. Au terme de cet exercice, il s'avère que l'effort cognitif et mnésique à mobiliser pour rapprocher les dispositifs cinématographiques augmente très fortement avec le nombre de films vus. Pour la majorité du panel, ce travail est ardu à partir d'une trentaine de films et semble impossible à mener, de manière rigoureusement identique, au-delà d'une centaine d'œuvres (soit pour la totalité du corpus de taille relativement modeste).

3.2 Analyse des réseaux de films

L'analyse de la recommandation du panel met en exergue une série de réseaux de films formalisés via un graphe réalisé avec l'outil Gephi (cf. figure 8). Seuls les films vus et recommandés entre eux par au moins 3 experts ont été pris en considération afin d'évacuer les liens faibles. Ce critère a permis d'identifier 5

groupes de films présentant une certaine autonomie et dont les partis pris de réalisation sont très proches²⁶.

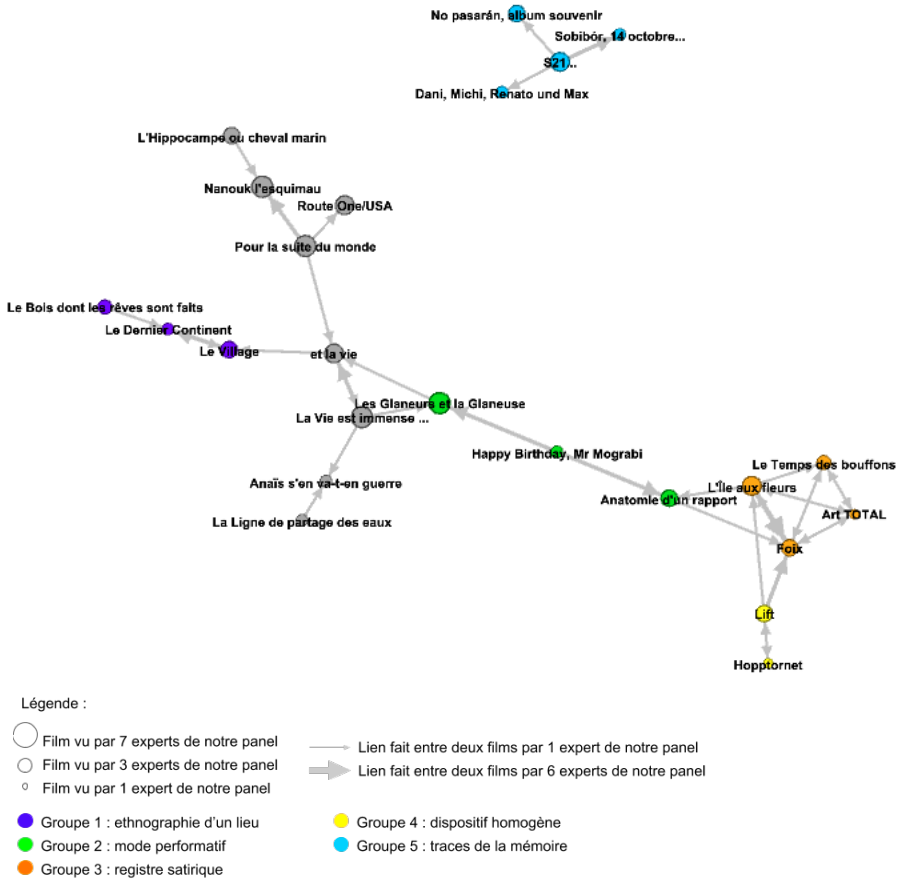


Figure 8. Graphe de réseaux de films identifiés par les experts²⁷

Le premier réseau de films dresse le portrait d'un lieu unique et des personnes qui l'habitent. Les décors et personnages sont filmés sur un temps long dans une démarche qui emprunte à l'observation ethnographique. Le deuxième groupe de films présente la démarche de réalisateurs qui se mettent en scène à l'image afin de provoquer l'action filmée. Leurs auteurs sont présents à la fois devant et derrière la caméra, et empruntent régulièrement à des dispositifs de mise en scène fictionnels.

²⁶ Le groupe constitué des films *Et la vie* et *La Vie est immense et pleine de danger* a été écarté de l'analyse en raison de son positionnement à l'intersection de plusieurs groupes avec lesquels il partage une partie des dispositifs de réalisation.

²⁷ Graphique généré sous Gephi au moyen des algorithmes de spatialisation Fruchterman Reingold puis Force Atlas 2, à partir des liens confirmés par 3 experts au moins entre films du corpus.

Le troisième groupe forme une clique de quatre films qui adoptent un registre satirique. Ces œuvres partagent une voix-off dont le contenu et le ton sont en décalage avec le montage image. Leurs propos évoquent la satire, voire un point de vue polémique sur différents sujets (écologique, néo-libéralisme, post-colonialisme, urbanisme). Un quatrième groupe rapproche deux films qui se caractérisent par une unité formelle s'appuyant à la fois sur un dispositif de prise de vue original et sur une cohérence spatiale et temporelle. Un cinquième groupe forme un ensemble séparé des autres films (situé en haut du graphe). Il interroge les traces laissées par un événement historique. Ces films mobilisent des archives, témoins, experts et adoptent une démarche de réalisation réflexive sur la construction mémorielle à laquelle ils participent.

3.3 Écarts entre recommandation humaine et algorithmique

Pour comprendre les différences entre recommandation humaine et algorithmique, l'analyse s'est focalisée sur les 5 groupes de films présentés ci-dessus (cf. figure 8). L'enjeu est de comparer les écarts entre les liens forts identifiés par les experts et les scores de proximité calculés par le moteur Rumo. À cette fin, chacun des films de ces 5 groupes ont été placés en regard les uns avec les autres dans la matrice d'adjacence ci-dessous. Chaque chiffre correspond à l'ordre d'apparition dans Rumo du film indiqué en tête de colonne parmi les recommandations associées au film en tête de ligne²⁸.

²⁸ Le moteur Rumo traduit le score calculé par la métrique cosinus en classant les 112 films du corpus, du plus proche au plus éloigné, selon la facette des dispositifs de réalisation. L'ordre croissant de proximité est indiqué du 1^{er} à 30^e rang avec un dégradé de couleur du rouge foncé au blanc.

Groupe	Titre des films	Le Bois...	Le Dernier...	Le Village	Anatomie...	Happy...	Les Glaneurs...	L' Île...	La Plage...	Le Temps...	Art Total	Foix	Hoptornet	Lift	Dani...	No pasarán...	S 21...	Sobibór...
Groupe 1 : ethnographie d'un lieu	Le Bois dont les rêves sont faits		10	11			25							14			20	26
	Le Dernier continent	12		10					9									
	Le Village	20	13										26	1			24	30
Groupe 2 : mode performatif	Anatomie d'un rapport				12	24		27			19	8	15		18			
	Happy Birthday, Mr Mograbi	28	22	9	4								5	7	29	19		
	Les Glaneurs et la glaneuse	27		30	2								22					
Groupe 3 : registre satirique	L'Île aux fleurs		18	30				17	2	5	3			28	27			
	La Plage des Shadocks			22			15	30	9	2	4	14						
	Le Temps des bouffons		5				2		17	1								
	Art Total		24				3	7	8		4							22
	Foix		30	24			2	7	1	8		21						
Groupe 4 : dispositif homogène	Hoptornet			17	8			13			27		1					
	Lift	12	1	14	9	16	18					2						17
Groupe 5 : traces de la mémoire	Dani, Michi, Renato und Max	25		30			28									2	1	3
	No pasarán, album souvenir			18	9	30							26	3		4	1	
	S 21, la machine de mort khmère rouge	13	10										29	1	4		2	
	Sobibór, 14 octobre 1943, 16 heures	22	14										18	5	1	4		

Tableau 1. Matrice d'adjacence algorithmique entre films étalons

Cette matrice d'adjacence met en exergue les liens forts établis par le moteur Rumo au sein des groupes de films. Les groupes 3 (registre satirique), 4 (dispositif homogène) et 5 (traces de la mémoire), qui représentent les films étalons les plus autonomes sur le graphe (figure 8), sont bien reliés entre eux par des liens forts. Quant aux groupes 1 (ethnographie d'un lieu) et 2 (mode performatif) qui occupent des positions intermédiaires dans le réseau de films, ils sont reliés entre eux mais avec des distances vectorielles supérieures ce qui explique des scores de proximité moins élevés. De manière globale, il semble possible d'établir une corrélation entre la recommandation experte et algorithmique basée sur le codage des dispositifs des films du corpus.

3.4 Discussion des résultats

Afin de comparer les apports et limites de cette preuve de concept, l'analyse s'est ensuite focalisée sur le groupe 3 (registre satirique). Les 2 figures ci-dessous présentent les réseaux de films en fonction des facettes thématiques ou dispositifs de réalisation ingérés dans le moteur Rumo pour cette clique.

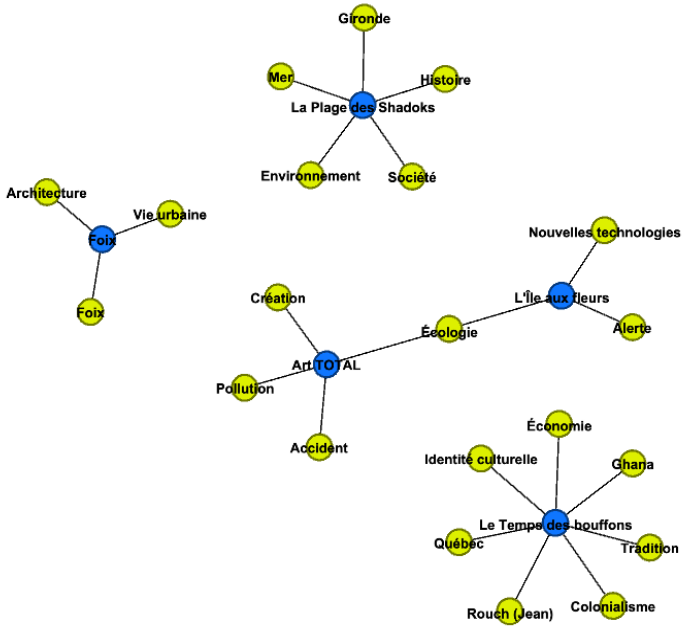


Figure 9. Liens entre films du groupe satirique via leurs descripteurs thématiques²⁹

²⁹ Les graphiques des figures 9 et 10 ont été générés avec le logiciel Gephi via l'algorithme Force Atlas 2. Le dé-chevauchement est manuel comme le rapprochement des cliques distinctes les unes des autres.

basée sur les dispositifs filmiques afin de renouveler les critères de classification utilisés dans la médiation du cinéma documentaire pour un public cinéphile.

Conclusion

L'évaluation de cette preuve de concept démontre l'intérêt de pouvoir qualifier le cinéma documentaire à travers un nouveau critère discriminant : ses dispositifs de réalisation. Dans un premier temps, des liens forts et totalement inédits entre des films issus du catalogue de Tënk ont pu être identifiés en utilisant la métrique Cosinus de Salton. Dans un deuxième temps, cette recommandation algorithmique a été validée par un panel d'experts cinéphiles et programmeurs du cinéma documentaire. Deux apports de connaissances ressortent de cette expérimentation exploratoire. Premièrement, ce dispositif de recommandation soulève l'intérêt d'équiper de manière alternative les plateformes de vidéo à la demande cinéphiles mais également toutes les institutions qui disposent de vastes fonds, collections ou catalogues de films documentaires. En effet, les métadonnées thématiques existantes à ce jour s'avèrent insuffisantes pour recommander le cinéma documentaire de manière totalement satisfaisante lorsqu'on s'intéresse aux dispositifs de réalisation. L'enjeu sous-jacent est de participer à la découvrabilité d'œuvres cinématographiques patrimoniales et contemporaines invisibilisées par la grande standardisation du marché télévisé ces vingt dernières années (Barreau-Brouste, 2011 ; Didier & Raynaud, 2014 ; Kilborne, 2011a, 2011b ; Lesaunier, 2019). Deuxièmement, cette recommandation automatisée permet de dépasser deux limites inhérentes à la programmation humaine : d'une part l'impossibilité pour un même expert de visionner la totalité des catalogues de plus en plus abondants et d'autre part l'effort cognitif et mnésique très important à fournir lorsque qu'un programmeur doit établir une analyse comparative des dispositifs de réalisation au-delà d'une cinquantaine de films.

Ce processus de conception a également permis d'identifier de nouvelles perspectives de recherche sur ce terrain d'étude. Premièrement, notre corpus – bien que représentatif du catalogue en termes de durée et de date – reste de petite taille. L'indexation de la totalité du catalogue de Tënk (environ 800 titres en 2020) avec le thésaurus de dispositifs de réalisation du cinéma documentaire devrait permettre d'observer empiriquement l'évolution des résultats algorithmiques : le volume du corpus a-t-il un impact sur la distance vectorielle calculée entre les films ? Observe-t-on des biais de catalogue lors du passage d'une centaine de films indexés à plusieurs milliers sur différents catalogues ? Deuxièmement, la robustesse du thésaurus doit être encore validée afin d'identifier l'ajout ou la suppression de descripteurs pendant le processus d'indexation. Une itération de ce jeu de données reste donc importante pour permettre de consolider la première version de cet outil en vocabulaire contrôlé. De plus, le moteur Rumo utilise, à ce jour, une liste de termes à plat. Cet outil ne permet donc pas de prendre en considération les liaisons hiérarchiques ou associatives entre les termes du thésaurus. Dès lors, il semble intéressant de se demander si la prise en compte de la structure hiérarchique ou associative entre les termes permettrait d'améliorer la pertinence des calculs de distance vectorielle. Troisièmement, la facette de données qualifiant les dispositifs de réalisation du cinéma documentaire pourrait être complétée par d'autres facettes de métadonnées (thématique, auteur, producteur, etc.). Il reste ainsi à démontrer si l'articulation de ces différentes facettes permettrait ou non d'améliorer la pertinence de cette recommandation automatisée. Finalement, cette recommandation

algorithmique devra être évaluée avec un panel d'utilisateur cinéphile de la plateforme afin de mieux appréhender dans quelle mesure et à quelle condition ce tri du catalogue peut les aider dans leur choix de visionnage.

Bibliographie

Ahn, S. & Shi, C.-K. (2009). Exploring Movie Recommendation System Using Cultural Metadata, in *Transactions on Edutainment II*, Pan, Z., Cheok, A. D., Müller, W., Rhalibi, A. E. (Eds.), Springer, Berlin, Heidelberg, 119-34.

Amatriain, X. (2013). Big & personal: data and models behind netflix recommendations, in KDD' 13: The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago Illinois –, *BigMine '13: Proceedings of the 2nd International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, Association for Computing Machinery, New York, 1-6.

Amer-Yahia, S., Favreau, A. & Sénéchal, J. (2020). *D'où vient le risque ? Des données et des algorithmes*. binaire (Le Monde.fr), Disponible à : <https://www.lemonde.fr/blog/binaire/2020/02/05/les-plateformes-numeriques-un-foyer-pour-les-risques-donnees-et-algorithmes/>.

Barreau-Brouste, S. (2011). *Arte et le documentaire : de nouveaux enjeux pour la création*. Bord de l'eau / INA, Lormont/Bry-sur-Marne.

Bernier, C., Nowakowski, S. & Boyer, A. (2011). Poursuite de cibles dans l'espace de recommandation - vers un nouveau système de recommandation basé sur le filtrage de Kalman, in *Conférence en Recherche d'Information et Applications - CORIA 2011*, Avignon, France, 57-65.

Bertail, P., Bounie, D., Cléménçon, S. & Waelbroeck, P. (2019). *Algorithmes : biais, discrimination et équité*. Télécom ParisTech, Fondation Aboena, 23 p., date last accessed janvier 27, 2021, sur <https://hal.telecom-paris.fr/hal-02077745>.

Beuscart, J.-S., Coavoux, S. & Maillard, S. (2019). Les algorithmes de recommandation musicale et l'autonomie de l'auditeur. *Réseaux*, num. 213, 17-47.

Bullich, V. (2018). Grandeur et décadence de l'« édition prédictive » : Retour sur l'intégration manquée d'une « machine prédictive » au sein d'une maison d'édition. *Réseaux*, num. 211, 257.

Burke, R. (2007). Hybrid Web Recommender Systems, in *The Adaptive Web: Methods and Strategies of Web Personalization*, Brusilovsky, P., Kobsa, A., Nejdl, W. (Eds.), Springer, Berlin, Heidelberg, 377-408.

Cardon, D. (2013). Dans l'esprit du PageRank. *Réseaux*, num. 177, 63-95.

Cardon, D. (2015). *À quoi rêvent les algorithmes. Nos vies à l'heure des big data*. Seuil, Paris.

Chapelain, B. & Ducas, S. (Eds.). (2018). *Prescription culturelle. Avatars et médiamorphoses*. ENSSIB, Villeurbanne.

- Citton, Y. (2014). *L'Économie de l'attention*. La Découverte, Paris.
- CPDirSIC. (2019). *Dynamiques des recherches en sciences de l'information et de la communication*. Conférence permanente des directeurs·trices des unités de recherche en sciences de l'information et de la communication, 194 p., date last accessed janvier 28, 2021, sur <https://hal.univ-lorraine.fr/hal-01885229>.
- Croissant, V. (Ed.). (2019). *L'Avis des autres. Prescription et recommandation culturelles à l'ère numérique*. éditions des archives contemporaines, Paris.
- Croissant, V. & Cambone, M. (2020). Discours et pratiques de la participation dans la construction des hiérarchies cinématographiques sur les plateformes de streaming vidéo, in *La Fabrique de la participation culturelle. Plateformes numériques et enjeux démocratiques –*, *La Fabrique de la participation culturelle. Plateformes numériques et enjeux démocratiques*, Severo, M. (Ed.), Paris, France, 57-66.
- CSA Lab. (2017). *Le rôle des données et des algorithmes dans l'accès aux contenus*. CSA. Les mutations de la mise à disposition de contenus audiovisuels à l'ère du numérique : conséquences et enjeux. 1, 20 p., date last accessed janvier 27, 2021, sur <https://www.csa.fr/Informer/CSA-lab/Les-publications/Les-mutations-de-la-mise-a-disposition-de-contenus-audiovisuels-a-l-ere-du-numerique-consequences-et-enjeux-Le-role-des-donnees-et-des-algorithmes-dans-l-acces-aux-contenus>.
- Dahimene, M. R. (2014). *Filtrage et Recommandation sur les réseaux sociaux*. Thèse de doctorat en informatique, CNAM.
- Debruyne, F. & Pirolli, F. (2017). Prescription et recommandation : agir et faire agir ? Introduction. *Études de communication*, num. 49, 7-12.
- Delaporte, C. (2019). La médiation générique des contenus cinématographiques sur les plateformes de vidéo à la demande. *Réseaux*, num. 217, 151-84.
- Didier, I. & Raynaud, P. (2014). L'extension des usages de l'archive audiovisuelle. *E-dossiers de l'audiovisuel*, Advance Access published juin 2014.
- Domenget, J.-C. & Coutant, A. (2014). Partir des usages pour analyser les systèmes de recommandation, in *Les Moteurs et systèmes de recommandation*, Kembellec, G., Saleh, I., Chartron, G. (Eds.), ISTE Editions, London, 43-67.
- Drumont, G. S. M., Countant, A. & Millerand, F. (2018). La production de l'utilisateur par les algorithmes de Netflix. *Les Enjeux de l'information et de la communication*, num. 19/2, 29-44.
- Ertzscheid, O. (2015). Usages de l'information numérique : comprendre les nouvelles enclosures algorithmiques pour mieux s'en libérer. *Revue française des sciences de l'information et de la communication*, num. 6, Advance Access published 2015: doi:10.4000/rfsic.1425.
- Farchy, J., Méadel, C. & Anciaux, A. (2017). Une question de comportement. Recommandation des contenus audiovisuels et transformations numériques. *Tic & société*, vol. 10, num. 2-3, 168-98.

Findeli, A. (2013 [2006]). Qu'appelle-t-on « théorie » en design ? Réflexions sur l'enseignement et la recherche en design, in *Le Design. Essais sur des théories et des pratiques*, Flamand, B. (Ed.), Éditions du Regard, Paris, 79-100.

Findeli, A. (2015). La recherche-projet en design et la question de la question de recherche : essai de clarification conceptuelle. *Sciences du Design*, num. 1, 45-57.

Foulonneau, M., Grouès, V., Naudet, Y. & Chevalier, M. (2014). Recommandeurs et diversité : Exploitation de la longue traîne et diversité des listes de recommandation, in *Les Moteurs et systèmes de recommandation*, Kembellec, G., Chartron, G., Saleh, I. (Eds.), ISTE Editions, London, 85-104.

Gantier, S. (2020). Construction de pratiques cinéphiles sur une plateforme de vidéo à la demande : enjeu du design des personae. *Les Enjeux de l'information et de la communication*, vol. 21, num. 1, 53-73.

Gantier, S., Bolka-Tabary, L., Kergosien, E., Givois, E. & Atbane-El Haoudi, B. (2020). *Rapport technique du programme AlgoDoc (Algorithme de recommandation de films documentaires) : guide d'indexation du thésaurus des procédés filmiques du cinéma documentaire*. Laboratoire GERüCO / Laboratoire DeVisu, 91 p., date last accessed janvier 27, 2021, sur <https://hal.archives-ouvertes.fr/hal-03041204>.

Gaudenzi, S. & Gantier, S. (2018). Vers un canevas méthodologique pour le design d'œuvres audiovisuelles interactives. *Interfaces numériques*, vol. 7, num. 3, 726-726.

Gauthier, G. (2015 [1995]). *Le Documentaire, un autre cinéma. Histoire et création*. Armand Colin, Paris.

Gentès, A. (2015). Arts et sciences du design : la place des sciences humaines. *Sciences du Design*, num. 1, 94-107.

Gentès, A. (2017). *The In-Discipline of Design: Bridging the Gap Between Humanities and Engineering*. Springer, Boston.

Gillespie, T. (2014). The Relevance of Algorithms, in *Media Technologies. Essays on Communication, Materiality, and Society*, Gillespie, T., Boczkowski, P. J., Foot, K. A. (Eds.), MIT Press, Cambridge, 167-94.

Givois, È., Bolka-Tabary, L., Kergosien, È., Gantier, S. & Atbane-El Haoudi, B. (2021). Élaboration itérative d'un thésaurus pour indexer le cinéma documentaire à partir des procédés filmiques. *I2D*, vol. 2.

Hachour, H., Szoniecky, S. & Abouad, S. (2014). Espaces sémiocognitifs : les frontières des systèmes de recommandation, in *Les Moteurs et systèmes de recommandation*, Kembellec, G., Chartron, G., Saleh, I. (Eds.), ISTE Editions Ltd, London, 161-90.

Jeanneret, Y. & Souchier, E. (1999). Pour une poétique de « l'écrit d'écran ». *Xoana*, num. 6, 97-107.

Kembellec, G., Chartron, G. & Saleh, I. (Eds.). (2014). *Les Moteurs et systèmes de recommandation*. ISTE Editions Ltd, London.

- Kilborne, Y. (2008). *L'Expérience documentaire : approche communicationnelle du cinéma de réalité*. Thèse de doctorat en sciences de l'information et de la communication dirigée par Jean-François Dégremont, Université Paris 8 Saint-Denis.
- Kilborne, Y. (2011a). Le petit écran, média indépassable ? Du statut de la télévision chez les cinéastes documentaristes. *Les Enjeux de l'information et de la communication*, num. 12/1, 95.
- Kilborne, Y. (2011b). Les stratégies de résistances des cinéastes documentaristes français. *Sociologie de l'art*, vol. 16, 81-98.
- Kirtchik, O. (2019). Qu'apprend-on des machines apprenantes ? *Zilse*, vol. 5, num. 1, 446-68.
- Kouadria, A. & Nouali, O. (2014). Amélioration du modèle de filtrage collaboratif hybride relatif au système de recommandation, in *JDI'2014*, Guelma, Algérie, 36-41.
- Labour, M. & Abou Jaoudé, C. (2014). Information, représentation et décision, in *Recherches en design. Processus de conception, écriture et représentations*, Leleu-Merviel, S., Boulekbache-Mazouz, H. (Eds.), ISTE Group, Londres, 149-203.
- Lallemand, C. & Gronier, G. (2018 [2015]). *Méthodes de design UX. 30 méthodes fondamentales pour concevoir des expériences optimales*. Eyrolles, Paris.
- Leleu-Merviel, S. (2008). *Objectiver l'humain. Qualification, quantification*. Lavoisier, Paris.
- Lemdani, R. (2016). *Système hybride d'adaptation dans les systèmes de recommandation*. Thèse de doctorat en sciences de la technologie et de la communication sous la direction de Yolaine Bourda, Université Paris-Saclay (ComUE).
- Lesauvier, M.-E. (2019). Une plateforme au service d'un monde professionnel mobilisé : enjeux symboliques et économiques. *Tic & société*, vol. 13, num. 1-2, 225-46.
- Lherisson, P.-R. (2018). *Système de recommandation équitable d'œuvres numériques. En quête de diversité*. Thèse de doctorat en informatique dirigée par Pierre Maret et Fabrice Muhlenbach, Université de Lyon.
- Linden, G. & Smith, B., York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, vol. 7, num. 1, 76-80.
- Lioult, J.-L. (2019). *À l'enseignement du réel : penser le documentaire*. Presses universitaires de Provence, Aix-en-Provence.
- Masure, A. (2017). *Design et humanités numériques*. Éditions B42, Paris.
- Ménard, M. (2014). Systèmes de recommandation de biens culturels. *Les Cahiers du numérique*, vol. 10, num. 1, 69-94.
- Musiani, F. (2013). Governance by algorithms. *Internet Policy Review*, vol. 2, num. 3.
- Nichols, B. (2010 [2001]). *Introduction to Documentary*. Indiana University Press, Bloomington.
- Niney, F. (2009). *Le Documentaire et ses faux-semblants*. Klincksieck, Paris.

- Paquienréguy, F. (2017). Le glissement de la prescription dans les plateformes de recommandation. *Études de communication. Langages, information, médiations*, num. 49, 13-32.
- Ricci, F., Rokach, L., Shapira, B. & Kantor, P. B. (2010). *Recommender Systems Handbook*. Springer, Boston.
- Rieder, B. (2020). *Engines of Order. A Mechanology of Algorithmic Techniques*. Amsterdam University Press, Amsterdam.
- Riondel, A. (2020). *Netflix : le binge-watching helvétique*. Mémoire de master en journalisme et communication, Université de Genève.
- Schmitt, É. (2016). Des humains dans la machine : la conception d'un algorithme de classification sémantique au prisme du concept d'objectivité. *Sciences du Design*, num. 4, 83-97.
- Sejwal, V. K. & Abulaish, M. (2021). CAMO: A context-aware movie ontology generated from LOD and movie databases. *Multimedia Tools and Applications*, vol. 80, num. 5, 7247-69.
- Taillibert, C. (2020). *Vidéo à la demande : une nouvelle médiation ? Réflexions autour des plateformes cinéphiles française*. L'Harmattan, Paris.
- Tardy, C. & Jeanneret, Y. (Eds.). (2007). *L'Écriture des médias informatisés. Espaces de pratiques*. Lavoisier, Paris.
- Thuillas, O. & Wiart, L. (2019). Les plateformes de VOD cinéphiliques. Des stratégies de niche en questions. *Les Enjeux de l'information et de la communication*, num. 20/1, 39-55.
- Vayre, J.-S. (2017). Une histoire de machines à recommander des biens de consommation. De l'Internet documentaire à l'Internet des données. *Études de communication. Langages, information, médiations*, num. 49, 89-106.
- Walek, B. & Fojtík, V. (2020). A hybrid recommender system for recommending relevant movies using an expert system. *Expert Systems with Applications*, vol. 158, 113452.
- Zacklad, M. (2017). *Design, conception, création vers une théorie interdisciplinaire du design*. Wikicréation, Disponible à : <https://wikicreation.fr/interdisciplinarite-et-creation/>.