



**HAL**  
open science

# État de l'art des méthodes de génération automatique de listes de lectures

Julien Aubert-Bédouchaud

► **To cite this version:**

Julien Aubert-Bédouchaud. État de l'art des méthodes de génération automatique de listes de lectures. 35èmes Journées d'Études sur la Parole (JEP) 31ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN) 26ème Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL), 2024, Toulouse, France. pp.42-56. hal-04622986

**HAL Id: hal-04622986**

**<https://inria.hal.science/hal-04622986v1>**

Submitted on 26 Jun 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.



Distributed under a Creative Commons Attribution 4.0 International License

# État de l'art des méthodes de génération automatique de listes de lectures

Julien Aubert-Bédouchaud<sup>1</sup>

(1) Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France  
julien.aubert-bedouchaud@univ-nantes.fr

## RÉSUMÉ

---

L'augmentation croissante du volume d'articles scientifique rend difficile la montée en compétence des chercheurs sur un domaine de recherche ciblé. Pour faciliter l'accès à ces articles, diverses approches et tâches en recherche d'information ont été développées ces dernières années. Parmi elles, la tâche de génération automatique de listes de lecture a été étudiée dans la littérature. Elle consiste en la génération d'une liste ordonnée d'articles scientifiques couvrant un domaine de recherche spécifique. Plusieurs travaux ont exploré différents aspects de cette tâche, proposant des jeux de données et des méthodologies d'évaluation variées pour apporter des solutions à ce problème. Dans cet article, nous présentons un état de l'art des principales approches de génération de listes de lecture, incluant les données, méthodes, et métriques d'évaluation.

## ABSTRACT

---

### **State-of-the-art of automatic reading lists generation methods**

The growing increase in the volume of scientific articles makes it difficult for researchers to improve their skills in a targeted area of research. To facilitate access to these articles, various information retrieval approaches and tasks have been developed. Among them, the task of automatic reading lists generation has been studied in the literature. It consists of the generation of an ordered list of scientific articles covering a specific area of research. Several works have explored different aspects of this task, proposing various datasets and evaluation methodologies to provide solutions to this problem. In this article, we present a state-of-the-art of the main reading lists generation approaches, including evaluation data, methods, and metrics.

---

**MOTS-CLÉS :** Listes de lecture, recherche d'information, recommandation d'articles, génération automatique.

**KEYWORDS:** Reading lists, information retrieval, article recommendation, automatic generation.

---

## 1 Introduction

La fréquence de publication des articles scientifiques est de plus en plus conséquente (Larsen & von Ins, 2010; Thelwall & Sud, 2022). L'impossibilité de lire tous les articles parus ainsi que le manque de moyen pour déterminer quels sont les articles les plus pertinents et utiles à un domaine rendent difficile la montée en compétence et le fait de garder ses connaissances à jour. Une manière d'avoir un accès rapide aux connaissances d'un domaine est de passer par les revues de littérature associées. Tous les domaines ne proposent néanmoins pas d'article de revue ou peuvent ne plus être à jour à mesure des publications du domaine. Des approches se concentrant sur les connaissances et

compétences attendues peuvent être privilégiées. Pour faciliter l'accès à ces informations, les experts d'un domaine peuvent proposer des listes de lecture (*reading lists* en anglais), listes de références académiques ordonnées permettant de couvrir de façon synthétique les notions importantes d'un domaine de recherche (Siddall & Rose, 2014). Un exemple de liste de lecture est proposé en Figure 1. En plus de permettre la montée en compétence d'un lecteur, ce type de liste permet de se positionner au sein d'un réseau d'articles scientifiques, les références à différents travaux permettant de tracer des liens entre différents aspects de la thématique recherchée.

A Discriminative Latent Variable Model for Statistical Machine Translation (*Blunsom et al., ACL 2008*)  
Probabilistic Inference for Machine Translation (*Blunsom & Osborne, EMNLP 2008*)  
A Hierarchical Phrase-Based Model for Statistical Machine Translation (*Chiang, ACL 2005*)  
What's in a translation rule? (*Galley et al., NAACL 2004*)  
11,001 New Features for Statistical Machine Translation (*Chiang et al., NAACL 2009*)

FIGURE 1 – Exemple de liste de lecture pour la thématique "statistical machine-translation models". (Jardine, 2014)

La génération de listes de lecture présentant un intérêt stratégique pour l'exploration d'un domaine scientifique, il est intéressant d'automatiser cette tâche. Soit un corpus d'articles scientifiques  $C = \{a_1, a_2, \dots, a_N\}$  et une requête  $R$  (mots-clés, articles de recherche, listes de notions, etc.), la tâche de génération de liste de lecture consiste à générer une liste ordonnée d'articles  $L$  telle que  $L \subset C$  où les articles de  $L$  sont sélectionnés en fonction de la requête  $R$ . Également, la liste de lecture doit permettre d'obtenir une vision large d'un sujet de recherche, le but étant de faire monter rapidement en compétence un chercheur sur un domaine. La liste de lecture peut s'apparenter à un graphe d'articles dont les arcs représentent les ordres de lecture inter-articles. Aussi, l'ordre des références présentes dans les listes de lecture à une importance, l'ordre de lecture pouvant être différent selon le niveau de l'utilisateur.

La tâche de génération automatique de liste de lecture est à la frontière de deux tâches du domaine de la recherche d'informations, mais présente des différences notables :

- *La recherche d'articles scientifiques* a pour objectif de trouver les documents répondant à un besoin utilisateur exprimé sous la forme d'une requête. L'utilisateur précise ainsi directement son besoin. (Kaur et al., 2010; Boudin et al., 2020).

Si cette approche propose des résultats ordonnés par pertinence, la liste de lecture doit également être ordonnée de façon à faciliter la compréhension d'une thématique.

- *La recommandation d'articles scientifiques* cherche à proposer des articles en fonction des intérêts d'un chercheur. Cette approche utilise une source de données initiale à partir de laquelle les recommandations sont déterminées selon la similarité entre l'utilisateur et les habitudes des autres utilisateurs. (Tang & Zhang, 2009; He et al., 2010; Beel et al., 2016).

La tâche de recommandation d'articles scientifiques peut proposer des documents similaires et redondants, ce qui n'est pas pertinent pour une liste de lecture.

La génération de liste de lecture est à la fois précise et synthétique (Thompson et al., 2004), cherchant un équilibre entre la pertinence des références et la couverture plus générale de la littérature existante. Une autre particularité de la liste de lecture académique est l'interchangeabilité des articles, les informations importantes d'un domaine pouvant être présentée par différents documents. La génération d'une liste de lecture a ainsi pour objectif de permettre une forme de navigation rapide au sein d'une

collection d'articles d'un domaine.

Cet article présente un état de l'art des travaux sur la génération de listes de lecture pour les collections d'articles scientifiques. La section 2 introduit les principales approches existantes, à commencer par les approches historiques issues du domaine de la recommandation, les modèles d'expertise et les approches utilisant de la modélisation de concepts. La section 3 s'intéresse aux jeux de données proposés pour cette tâche, à savoir les graphes de citations et les listes de référence. La section 4 présente les différentes métriques utilisées lors de l'évaluation des systèmes proposés ainsi qu'un aperçu des méthodes d'évaluation utilisateur mises en place. S'ensuivra une discussion (Section 5) quant aux perspectives de ce domaine de recherche ainsi qu'une conclusion.

## 2 Approches existantes

Plusieurs façons d'approcher la tâche de génération automatique de listes de lecture ont été explorées dans la littérature. Les approches historiques (Section 2.1) se sont initialement intéressées à la recommandation d'articles fondateurs (c'est-à-dire les articles initiaux ayant contribué de façon significative au domaine) par filtrage collaboratif (Tang, 2008; Ekstrand *et al.*, 2010). Ces méthodes nécessitent l'évaluation d'articles de la part d'autres utilisateurs, qu'elle soit directe (notes utilisateurs) ou indirecte (nombre de citations d'un article).

D'autres approches ont cherché à modéliser l'importance des articles à travers des notions marquant l'expertise d'un domaine (Section 2.2) en s'appuyant sur des informations liées aux articles (Jardine, 2014; Sesagiri Raamkumar *et al.*, 2017; Figueira *et al.*, 2019).

Enfin, une troisième famille d'approches s'intéresse aux dépendances nécessaires à la compréhension de certains sujets (Section 2.3). Ces méthodes s'intéressent notamment aux relations entre les concepts présents dans une collection d'articles dans le but de déterminer des chaînes de prérequis (Gordon *et al.*, 2017). Ces travaux ont conduit Ding *et al.* (2022) à s'intéresser à la construction de chemins de lecture à partir des concepts implicites d'un graphe de citation.

### 2.1 Approches liées à la recommandation d'articles scientifiques

Les approches historiques liées à la tâche de génération de listes de lecture sont issues du domaine de la recommandation. Plusieurs approches de recommandation traditionnelles ont été adaptées à la tâche, à savoir le filtrage collaboratif proposant d'utiliser les avis d'autres utilisateurs pour une recommandation de documents pertinents, le filtrage basé sur le contenu reposant sur la similarité entre les données des documents, ou encore les filtrages hybrides combinant les prédictions des techniques précédentes. Ces techniques sont utilisées à partir d'évaluation manuelle des articles de la collection (Section 2.1.1) ou par l'utilisation des citations inter-articles (Section 2.1.2).

#### 2.1.1 Recommandation par notes utilisateurs

Tang (2008) a considéré la tâche de génération de liste de lecture comme une extension des tâches de recommandation traditionnelles. La génération de listes se distinguait ici par une recommandation plus large que les préférences du lecteur, les notions nécessaires à la compréhension d'un cours

ayant un impact sur les articles à recommander. Différentes techniques de filtrage ont été mises au point. À partir de notations d'articles de la part d'utilisateur du système, différents algorithmes cherchaient à répondre à des scénarios d'apprentissage (étudiant débutant un cours, se situant à la moitié, à la fin, nombre d'articles ou de notations suffisants ou non). Une approche par filtrage hybride, *PopCon2D*, fut identifiée comme répondant le mieux aux différents scénarios d'avancée du lecteur et de disponibilité de notations. Ce filtrage sélectionnait les articles à partir de la note moyenne, de la similarité entre les thèmes de l'article et ceux du champ d'étude que l'utilisateur avait au préalable déclaré connaître ainsi que les notes des utilisateurs ayant des habitudes semblables à l'utilisateur actif.

## 2.1.2 Recommandation par citations et co-références

Ekstrand *et al.* (2010) ont constaté que les techniques classiques de filtrage collaboratif utilisées dans le contexte d'articles scientifiques ne permettent que l'exploitation du contexte local d'un article, n'exploitant en effet que les coréférences d'articles et pas les relations ayant lieu au sein d'un domaine.

À partir d'un échantillon de documents d'un domaine, des techniques d'ordonnement de sommets de graphe permettaient de mesurer l'importance des articles de l'échantillon au sein d'une collection plus large. L'impact individuel et l'influence sur d'autres travaux furent mesurés par *PageRank* (Page *et al.*, 1998), *HITS* (Kleinberg, 1999), *SALSA* (Lempel & Moran, 2001), *HITS with Priors* (White & Smyth, 2003) et *k-step Markov* (White & Smyth, 2003). La mesure d'importance obtenue était couplée à différentes approches de filtrage dérivées des modèles de recommandation. Parmi elles, l'approche privilégiée était un algorithme de filtrage collaboratif utilisant les citation inter-articles pour déterminer le contexte local et *PageRank* pour mesurer l'importance de l'article au sein du domaine.

## 2.2 Modélisation de l'expertise du domaine

Les techniques de recommandation étaient limitées par la nécessité de connaître les préférences de la communauté de chercheurs dans le processus de filtrage. Ces informations pouvant ne pas être disponibles, des travaux ont cherché à mesurer l'importance des articles en modélisant l'expertise d'un domaine à travers les thématiques sous-jacentes des articles (Section 2.2.1), les mots-clés spécifiés par les auteurs (Section 2.2.2), la probabilité de citation de la part d'un lecteur (Section 2.2.3) ou une agrégation de plusieurs critères (Section 2.2.4).

### 2.2.1 Couverture de thématiques latentes

Jardine (2014) a cherché à quantifier l'importance des articles selon leur correspondance aux thématiques scientifiques présentes dans une collection d'articles scientifiques. Dans ce but, une modification de l'algorithme *PageRank*, *ThemedPageRank* (TPR), était proposée par Jardine & Teufel (2014). Le processus identifiait tout d'abord un sac-de-termes-techniques à partir des titres d'articles, puis une phase de modélisation des thématiques sous-jacentes du corpus était réalisée. La modélisation de thématique pouvait être réalisée par allocation latente de Dirichlet (LDA) (Blei *et al.*, 2003) ou factorisation de matrice non-négative (NMF) (Lee & Seung, 2000). L'étape de classement de sommets de graphes de TPR permettait ensuite de quantifier l'expertises des articles sur chacune des

thématiques identifiées.

## 2.2.2 Couverture de thématiques explicites

Sesagiri Raamkumar *et al.* (2017) ont mis au point une mesure de couverture thématique et périphérique (*Topical and peripheral coverage*) (TPC). Celle-ci s'appuyait sur les mots-clés spécifiés par l'auteur d'un article dans le but d'identifier un ensemble d'articles possédant des mots-clés similaires et ayant un lien de citation avec l'article initial. Cette métrique cherchait à couvrir un ensemble de critères de contenu présents dans la liste : les articles *populaires ou fondateurs*, articles de *revue de littérature*, articles *récents* et articles *associés à un sous-domaine de recherche*.

La mesure de TPC fut utilisée au sein d'un processus de recherche basée sur les mots-clés spécifiés par l'auteur (*Author-specified Keywords based Retrieval*) (AKR). Le processus identifiait une liste d'articles similaires au sujet de recherche via Okapi BM25 (Robertson *et al.*, 1996) puis triait les résultats en fonction de la somme pondérée du nombre de citations, de références et du score de TPC d'un article. La liste de lecture obtenue était ainsi la liste des articles contribuant de façon importante dans leurs thématiques respectives, en termes de citations ou de références.

## 2.2.3 Probabilité de citation d'un article

Figueira *et al.* (2019) ont étendu les travaux de Sesagiri Raamkumar *et al.* (2017) en proposant un critère supplémentaire caractérisant l'intérêt des listes de lecture : la présence d'*articles ayant une forte probabilité d'être cité* par l'utilisateur de la liste de lecture.

Des méthodes par apprentissage supervisé furent proposées dans le but de prédire des classements de listes d'articles optimisés sur le critère de citation. Des méthodes *Learning to Rank* (Liu, 2009) permettant d'apprendre une fonction de classement à partir de données existantes ont ainsi été mises en place. Plusieurs méthodes furent exploitées pour cette approche, à savoir MART (*Multiple Additive Regression Trees*) (Friedman, 2001), *LambdaMART* (Burgess, 2010), *ListNet* (Cao *et al.*, 2007), *RankNet* (Burgess *et al.*, 2005), *AdaRank* (Xu & Li, 2007), *RankBoost* (Freund *et al.*, 2003), des forêts d'arbres décisionnels (Ho, 1995) et algorithmes d'ascension de coordonnées (Fessler *et al.*, 1997). Les modèles étaient entraînés avec des caractéristiques (*features*) tels que le score de TPC (Sesagiri Raamkumar *et al.*, 2017), l'âge des articles, le fait qu'un article soit un article de revue de littérature ou non, le nombre de références des articles associés aux mots-clés recherchés, des mesures de pondération de documents telles que BM25 ainsi qu'une mesure de similarité cosinus.

## 2.2.4 Agrégation de multiples critères

Figueira *et al.* (2019) ont également proposé des méthodes d'agrégation de classement pour la génération de listes de lecture, permettant de classer des listes sur la base des critères définis dans les sections 2.2.2 et 2.2.3.

Une phase de génération d'une liste d'articles en lien avec la requête était effectuée via Okapi BM25. La phase de classement était ensuite réalisée au moyen de différentes techniques d'agrégation de rangs : méthode Borda (de Borda, 1781), des techniques de comptage binaire ou classement par maximisation du gain cumulé. Les prédictions des méthodes *Learning to Rank* (Section 2.2.4) furent

utilisées dans certaines variantes afin d’estimer la probabilité d’un article d’être cité par le lecteur. Les scores de classement produits par LambdaMART furent utilisés pour estimer cette probabilité.

## 2.3 Utilisation de chaînes de prérequis

Les listes de lectures générées par les approches précédentes proposent un ordre de lecture dépendant de son impact sur une communauté d’utilisateurs (Section 2.1) ou de la correspondance à un critère spécifique (Section 2.2). Des travaux plus récents se sont intéressés à la construction d’un ordre de lecture s’appuyant sur les *concepts prérequis* pour la compréhension de ressources pédagogiques (Gordon *et al.*, 2016; Liu *et al.*, 2016; Pan *et al.*, 2017). Des approches se sont concentrées plus spécifiquement sur le domaine des articles scientifiques, cherchant à expliciter les concepts latents (Section 2.3.1) ou les induire par co-citations (Section 2.3.2).

### 2.3.1 Explicitation de concepts latents

Gordon *et al.* (2017) se sont intéressés à l’utilisation des *concepts*, notions nécessaire à la compréhension d’une notion pédagogique, pour la génération de listes de lecture. Cette approche proposait d’identifier les concepts présents dans les articles à partir des thématiques extraites d’un corpus par LDA à l’instar de Jardine & Teufel (2014). Des relations inter-concepts étaient ensuite calculées par similarité de Jaccard et méthode d’entropie croisée, l’association d’articles à des concepts permettant de déterminer des dépendances au sein d’une collection de documents scientifiques. La liste de lecture était ensuite générée en fonction de l’intérêt mesuré par chevauchement de lexique entre concepts et requête utilisateur. Les articles associés à un concept étaient identifiés par un parcours en profondeur du graphe. Cette approche calculait également un ordre de lecture via le niveau de familiarité de l’utilisateur avec les concepts des documents en fonction des réponses à un questionnaire ou des documents lus par l’utilisateur.

### 2.3.2 Identification de concepts par co-citations

Ding *et al.* (2022) se sont intéressés à la façon de trouver un chemin de lecture au sein d’un graphe de citations, à savoir un ordre de lecture spécifique dépendant des concepts pré-requis à la compréhension d’une thématique à partir des relations de co-citation des articles. Cette approche, proche de la génération de liste de lecture, s’intéressait à la génération d’une liste de références pertinente pour une requête contenant également les articles pré-requis aux références identifiées.

L’approche de Ding *et al.* (2022) proposait l’ordonnement d’une liste d’articles scientifiques par relation de dépendance inter-articles, à partir d’une requête constituée de mots-clés. Une liste initiale d’articles était générée par Google Scholar<sup>1</sup> à partir des mots-clés d’une requête. Cette liste initiale, couplée à un graphe de citations de six millions d’articles du domaine de l’informatique, permettait de construire un sous-graphe de citations incluant les articles récupérés ainsi que leurs voisins proches. Les concepts prérequis étaient capturés implicitement à partir des informations de co-citation des articles de la liste initiale présents dans le graphe de citation permettant de détecter de nouveaux articles nécessaires à la compréhension. L’algorithme *Node-Edge Weighted Steiner Tree* (NEWST) (Segev, 1987) permettait de détecter un arbre couvrant minimal du sous-graphe pour

---

1. <http://scholar.google.com>

généraliser un chemin de lecture optimal. Le sens de lecture de la liste était ensuite déduit à partir des coréférences et dates de publication d’articles.

### 3 Jeux de données

Différents jeux de données ont été construits et utilisés pour les besoins de la tâche de génération de liste de lecture. Deux types d’ensembles se distinguent dans les travaux réalisés. Des collections d’articles scientifiques (Table 1) couvrent les articles d’un ou plusieurs domaines scientifiques et peuvent être utilisés pour la conception de graphes de citations, l’apprentissage, ou l’évaluation des modèles. Des corpus de référence (Table 2) permettent, quant à eux, l’évaluation des listes générées.

#### 3.1 Collections d’articles scientifiques

Article	Source	Domaine	Plage	Nb Documents	Langue	Citations	Disponible
Tang (2008)	IEEE et ACM	Ingénierie informatique	1992-2004	21	en	✗	✓
Ekstrand <i>et al.</i> (2010)	ACM DL	Informatique	N.C.	201 145	en	✓	✗
Jardine (2014)	ANN	Traitement automatique des langues (TAL)	1960s - 2010	15 388	en	✓	✗
Gordon <i>et al.</i> (2017)	ANN, Wikipédia, ScienceDirect, Tutoriel Web	TAL	1960s - 2014	25 319	en	✓	✗
Sesagiri Raamkumar <i>et al.</i> (2017)	ACM DL	Informatique	1950-2011	103 739	en	✓	✗
Figueira <i>et al.</i> (2019)	Scopus	Informatique, Ingénierie	1970-2018	58 734	en	✓	✗
Ding <i>et al.</i> (2022)	S2ORC	Informatique	1913-2020	6 000 000	en	✓	✗

TABLE 1 – Résumé des collections d’articles scientifiques utilisées.

Les collections d’articles scientifiques présentés dans la littérature peuvent être des ensembles d’articles ou des réseaux de citations. Différentes sources ont été utilisées, telles l’*ACM Digital Library* (Ekstrand *et al.*, 2010; Sesagiri Raamkumar *et al.*, 2017), l’*ACL Anthology Network* (AAN) (Jardine, 2014; Gordon *et al.*, 2017), des journaux IEEE et ACM (Tang, 2008), Scopus (Figueira *et al.*, 2019) ou le corpus de *Semantic Scholar* (S2ORC) (Ding *et al.*, 2022). Si les collections utilisés sont de taille variable selon les approches, allant de la dizaine d’articles à 6 millions de documents, les tailles sont généralement de l’ordre de la dizaine à la centaine de milliers d’articles. Malgré différentes sources de données, les documents portent tous sur le domaine de l’informatique et présentent des articles scientifiques en langue anglaise seulement. Parmi les collections utilisées dans la littérature, seul celle de Tang (2008) est disponible, en annexe de son manuscrit de thèse. Les autres articles mentionnés proposent des indications permettant de construire des collections semblables à celles utilisées, mais ne sont pas reproductibles à l’identique.

#### 3.2 Corpus de référence

Les corpus de références sont utilisés à des fins d’évaluation de la tâche de génération automatique de listes de lecture. Ekstrand *et al.* (2010) ont utilisé un jeu de données de référence constitué à partir de 220 articles de l’*ACM Computing Survey* présentant au moins 15 références. Les données utilisées ne sont néanmoins pas disponibles. Jardine (2014) a proposé un corpus ayant été construit par des experts en TAL à partir de l’AAN. Huit listes de lecture y sont proposées, couvrant différentes



Article	Source	Domaine	Plage	Nb Documents	Nombre de références	Langue	Disponible
Ekstrand <i>et al.</i> (2010)	ACM Computing Surveys	Informatique	N.C.	220	15+	en	✗
Jardine (2014)	ANN (Annotations expertes)	TAL	1991-2009	8	~12	en	✓
Figueira <i>et al.</i> (2019)	Scopus (Articles classifiés <i>Reviews</i> et <i>Short Surveys</i> )	Informatique, Ingénierie	1970-2018	1648	N.C.	en	✗
SurveyBank (Ding <i>et al.</i> , 2022)	S2ORC, Google Scholar	TAL, Machine Learning, Intelligence Artificielle, Apprentissage profond, Recherche d'informations	1913-2020	9 321	~58	en	✓

TABLE 2 – Résumé des corpus utilisés comme référence aux listes de lecture.

thématiques de recherche. Ce jeu de données était utilisé par Gordon *et al.* (2017) pour une évaluation utilisateur. Parmi la collection de Figueira *et al.* (2019), 1648 articles sont considérés comme articles de revue, permettant d'utiliser leur listes de citations comme ensemble de références. Ding *et al.* (2022) ont proposé une collection d'articles de revue, SurveyBank<sup>2</sup>. Le jeu de données est constituée à partir de S2ORC et Google Scholar en utilisant les mots-clés du domaine de l'informatique issus de TutorialBank (Fabbri *et al.*, 2018) et LectureBank (Li *et al.*, 2019). Cette collection est initialement pensée pour la tâche de génération de chemin de lecture, mais peut être utilisée comme jeu de références pour la tâche de génération de listes de lecture.

## 4 Méthodes d'évaluation

### 4.1 Comparaison de la liste générée à une référence

Différentes métriques d'évaluation de la capacité des modèles à générer une liste de lecture à partir d'une collection ont été proposées dans la littérature. Ces métriques cherchaient à comparer une liste générée à une liste de référence sur la base de différents critères. Un résumé des méthodes de comparaison de la liste générée à une référence est proposé dans la Table 3.

Ekstrand *et al.* (2010) ont cherché à mesurer parmi les algorithmes mis en place ceux présentant l'ordre le plus utile pour les utilisateurs, utilisant des données issues d'articles de *survey* comme référence. La métrique d'utilité de la demi-vie (*half-life utility metric*) (Breese *et al.*, 1998) permettait d'estimer la probabilité qu'un utilisateur ait besoin d'un article d'une liste ordonnée. L'utilité attendue de la liste de lecture générée se mesurait ainsi en tenant compte de la position des articles générés présents dans la liste de référence. L'utilité était pondérée en fonction du rang de l'article : plus l'article est généré tard dans la liste, moins son score est élevé. Cette métrique possédait le défaut de ne pas permettre la mesure des articles pouvant se substituer à d'autres, l'utilité étant définie par la présence ou non d'un article dans la liste générée.

Jardine (2014) a évalué la tâche sur un ensemble de méthodes historiques, et des métriques visant à combler leurs lacunes. Les métriques historiques de **précision**, **rappel** et **F1-score** (Rijsbergen, 1979) ainsi que la moyenne des précisions (*Mean Average Precision*) (MAP) ont été employées, mais ne permettaient pas de prendre en compte les documents interchangeable. Ding *et al.* (2022) ont également utilisé des mesures de **précision@K** et **F1-score@K** (Manning *et al.*, 2008) dans le but de comparer les listes de lecture générées par leur modèle NEWST aux résultats de plateformes de recherche d'articles en utilisant les références de leur jeu de données SurveyBank.

2. <https://github.com/JiayuanDing100/Reading-Path-Generation>

Métriques	Intuition	Avantages	Inconvénients
<i>Précision, Rappel, F1score</i>	Mesures de la pertinence des articles	✓ Métriques historiques	✗ Correspondance partielle non considérée ✗ Ne prends pas en compte l'ordre des documents
<i>Half-life utility metric</i>	Estimation de l'utilité des articles d'une liste pour un utilisateur	✓ Prise en compte de l'ordre des documents	✗ Articles substituables non considérés
<i>MAP</i>	Précision moyenne des éléments d'une liste pour un ensemble de requêtes	✓ Prise en compte de l'ordre des documents	✗ Correspondance partielle non considérée
<i>RCP</i>	Rapport entre le nombre d'articles citant à la fois les articles générés et les articles de référence et le nombre d'articles citant les articles de référence	✓ Prise en compte des correspondances partielles	✗ Impossibilité de mesurer les articles non-cités ✗ Pertinence partielle trop importante dans la mesure du score
<i>CSC</i>	Mesurer la distance de citation entre articles d'une liste générés et celle de référence	✓ Prise en compte des correspondances partielles ✓ Prise en compte des articles non-cités	✗ La co-citation seule ne permet pas de détecter la similarité

TABLE 3 – Résumé de métriques de comparaison de la liste générée à une référence.

Autre méthode explorée par [Jardine \(2014\)](#), la métrique de probabilité relative de co-citation (**Relative co-cited probability**) (**RCP**) ([He et al., 2011](#)) permettait de déterminer des articles alternatifs à un article de référence cité par un auteur en mesurant le nombre de fois où deux articles (référence et générés) ont été co-cités dans un corpus par rapport au nombre de citations de la référence seule. Si cette métrique était utilisée par [Jardine \(2014\)](#) à titre d'information, un désavantage de cette méthode était l'impossibilité de mesurer la probabilité des articles n'ayant pas de citations, bien que le contenu puisse être similaire. Également, les articles référencés par plusieurs articles de la collection avaient un impact plus important dans l'équation, étant parfois plusieurs fois pris en compte.

Le coefficient de substitution de citation (**Citation Substitution Coefficient**) (**CSC**) était une métrique proposée par [Jardine \(2014\)](#) estimant la pertinence partielle entre deux articles d'un graphe de citations. Le nombre minimal de nœuds nécessaire pour relier deux articles au sein d'un graphe de citations permettait d'évaluer la distance de citations entre une liste d'articles de référence et une liste proposée par un système de génération de listes de lecture. Par rapport à la mesure de RCP, le CSC considérait les articles peu cités ayant des références aux articles pertinents pour le domaine. Une limite de cette approche résidait dans le fait que la co-citation d'articles ne marque pas systématiquement la similarité ou l'interchangeabilité.

## 4.2 Évaluation de critères

Un autre pan de recherche dans le domaine cherchait à évaluer la capacité des systèmes de génération de listes de lecture à répondre à un ensemble de critères définis dans le but de déterminer une liste pertinente. La liste des critères concernés est couverte par les sections [2.2.2](#) et [2.2.4](#).

La **méthode de l'entropie croisée** était proposée par [Sesagiri Raamkumar et al. \(2017\)](#) afin d'évaluer le système AKR. L'approche utilisée impliquait l'agrégation de classements calculés sur la base des prérequis proposés par l'article, la méthode satisfaisant le plus de prérequis étant celle sélectionnée.

Deux distances furent utilisées pour cette méthode, le  $\rho$  (rho) de Spearman (Spearman, 1904) mesurant la distance de déplacements entre deux éléments permutés et le  $\tau$  (tau) de Kendall (Kendall, 1938) mesurant le nombre d'inversions dans un classement (Kumar & Vassilvitskii, 2010).

Figueira *et al.* (2019) ont utilisé la métrique de **précision@K**, afin d'évaluer le pourcentage d'articles pertinents dans un ensemble de candidats par rapport aux prérequis proposés par l'auteur. La mesure était utilisée pour évaluer la popularité, la présence d'articles de revue de littérature, la présence d'articles récents et la probabilité de citation. La diversité des listes de lecture était mesuré au moyen d'une métrique d' $\alpha$ -NDCG@20 (Clarke *et al.*, 2008) (Vargas, 2014).

### 4.3 Évaluation humaine par les utilisateurs

Article	Mode d'évaluation	Critère évalué	Nb Participants	Conclusion de l'étude
Tang (2008)	Questionnaire à échelle d'évaluation	Évaluation de la qualité des articles proposés	25	Les approches utilisant un filtrage hybride à partir de la notre moyenne, de la similarité entre les thèmes de l'article et ceux du champ d'étude de l'utilisateur, ainsi que les notes des utilisateurs ayant des habitudes semblables à l'utilisateur actif permettent de bons résultats sur les scénarios de recommandation.
Ekstrand et al. (2010)	Questionnaire à échelle d'évaluation	Évaluation de la qualité des articles proposés	19	Les utilisateurs trouvent le système développé prometteur.
Sesagiri Raamkumar et al. (2017)	Questionnaire à échelle d'évaluation	Évaluation de la pertinence de la liste en fonction de critères spécifiques	132	La présence d'articles fondateurs a un impact direct sur la satisfaction des participants Les étudiants ont plus d'intérêt pour le système que les chercheurs confirmés
Gordon et al. (2017)	Questionnaire à échelle d'évaluation	Évaluation du potentiel pédagogique de la liste générée	33	Les listes générées présentent un potentiel pédagogique allant de modéré à élevé selon les domaines.
Gordon et al. (2017)	Modifications de listes générées et de références	Calcul de la distance de Levenshtein entre les listes modifiées et les listes proposées	33	Les listes générées sont comparables aux listes expertes.
Ding et al. (2022)	Questionnaire à choix multiples	Évaluation de la pertinence, exhaustivité et de la présence de dépendances dans les listes générées	16	L'utilisation de chaînes de prérequis augmente la connaissance apportée par les listes générées

TABLE 4 – Résumé des méthodes d'évaluation utilisateur.

Il est difficile de tirer des conclusions sur l'utilisation des systèmes par un humain uniquement par des méthodes d'évaluation automatisées. Plusieurs types d'évaluations utilisateur ont ainsi pu être mis en place dans les travaux réalisés (Table 4).

Gordon *et al.* (2017) ont évalué les performances du système sur une expérience intégrant des utilisateurs. L'expérimentation implique 33 chercheurs en TAL chargés de modifier des listes générées ainsi que de référence afin qu'elles atteignent un niveau convenable selon eux. L'expérience cherche à comparer la liste originale et la correction au moyen de la distance de Levenshtein (Levenshtein, 1966) afin de comparer le niveau de modification entre les deux. Les résultats obtenus indiquent que les listes générées sont comparables aux listes expertes

Des questionnaire à échelle d'évaluation (Likert, 1932) ont été proposés par plusieurs travaux. Tang (2008) a demandé à des 25 étudiants de noter les articles d'une liste afin de mesurer la pertinence des algorithmes de filtrage dans différents scénarios (nombre d'articles, avancement dans le cours). Ekstrand *et al.* (2010) ont posé des questions à 19 novices sur leur familiarité avec les articles de la liste générée, la pertinence des articles sélectionnés pour une thématique et l'importance des thématiques traitées. Les résultats indiquent que les utilisateurs trouvent le système prometteur. (Gordon *et al.*, 2017) ont évalué leur système sur la pertinence perçue de 33 chercheurs en TAL, indiquant un potentiel pédagogique modéré ou élevé selon les listes générées. Sesagiri Raamkumar

*et al.* (2017) ont évalué des listes générées en fonction des domaines de recherche de 132 participants (62 étudiants, 70 chercheurs), sur la pertinence de la liste en fonction de différents critères (familiarité, utilité, etc.). Les résultats indiquent que la présence d'articles fondateurs a un impact direct sur la satisfaction et que les étudiants ont plus d'intérêt pour le système que les chercheurs confirmés.

*Ding et al.* (2022) ont utilisé un questionnaire à choix multiple pour comparer le système NEWST à Google Scholar. 16 diplômés évaluent la pertinence des listes générées, leur exhaustivité et si les articles prérequis sont identifiables. Les résultats indiquent que les chaînes de prérequis augmentent la connaissance apportée par les listes générées.

## 5 Discussion

Les différentes approches de la littérature ont proposé des interprétations différentes des besoins de la tâche de génération de listes de lecture. Les approches historiques ont ainsi développé principalement des méthodes de recommandation améliorées par des techniques d'ordonnement de sommets de graphe. Des approches plus modernes tentent de modéliser l'expertise du domaine pour l'ordonnement des articles, d'autres explorent la modélisation de dépendances d'articles à des concepts. Ces différentes approches s'intéressent peu au contenu et informations sous-jacentes des articles scientifiques, des méthodes plus modernes de traitement automatique des langues pourraient permettre l'émergence de nouvelles méthodes de génération de listes de lecture. Certaines facettes de la tâche restent également peu explorées, notamment l'ordre de lecture des articles ou la minimisation de la redondance entre les articles retournés. L'introduction de mesure de sérendipité sur cette tâche (mesurant dans ce cas le fait de découvrir de façon fortuite un article pertinent) pourrait également être exploré afin de permettre la construction de listes favorisant les démarches de recherche scientifique.

Parmi les collections présentées, la majorité des corpus ne sont pas mis à disposition ou ne proposent pas un nombre d'échantillons suffisamment conséquent pour entraîner un modèle. Ce manque de données explique la quantité limitée d'approches supervisées et l'absence de méthodes neuronales dans le traitement de la tâche de génération de listes de lecture. Les avancées liées à la similarité sémantique pourraient notamment être exploitées pour la génération de listes de lecture. Les représentations vectorielles du contenu pourraient entre autres permettre de détecter plus efficacement les articles substituables d'une collection.

L'évaluation des systèmes de génération de listes de lecture a été réalisé de différentes façons dans la littérature. Néanmoins, le manque de méthodes d'évaluation couvrant plusieurs approches et les données non comparables ne permettent pas de comparer de façon exhaustive les travaux associés à cette tâche. L'évaluation de listes de lectures générées automatiquement reste donc un défi auxquels les futurs travaux devront s'intéresser.

## Remerciements

Je tiens à remercier mon équipe encadrante, Florian Boudin, Richard Dufour et Béatrice Daille, pour leurs conseils et leur soutien. Merci également aux membres de l'équipe TALN du LS2N et aux relecteurs RECITAL pour leurs commentaires et suggestions. Ce travail est financé dans cadre du projet AID-CNRS NaviTerm (convention 2022 65 0079 CNRS Occitanie Ouest).

# Références

- BEEL J., GIPP B., LANGER S. & BREITINGER C. (2016). Research-paper recommender systems : a literature survey. *International Journal on Digital Libraries*, **17**, 305–338. DOI : [10.1007/s00799-015-0156-0](https://doi.org/10.1007/s00799-015-0156-0).
- BLEI D. M., NG A. Y. & JORDAN M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, **3**, 993–1022.
- BOUDIN F., GALLINA Y. & AIZAWA A. (2020). Keyphrase generation for scientific document retrieval. In D. JURAFSKY, J. CHAI, N. SCHLUTER & J. TETREAU, Éd., *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, p. 1118–1126, Online : Association for Computational Linguistics. DOI : [10.18653/v1/2020.acl-main.105](https://doi.org/10.18653/v1/2020.acl-main.105).
- BREESE J. S., HECKERMAN D. & KADIE C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, UAI'98, p. 43–52, San Francisco, CA, USA : Morgan Kaufmann Publishers Inc.
- BURGES C., SHAKED T., RENSHAW E., LAZIER A., DEEDS M., HAMILTON N. & HULLENDER G. (2005). Learning to rank using gradient descent. In *Proceedings of the 22nd International Conference on Machine Learning*, ICML '05, p. 89–96, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/1102351.1102363](https://doi.org/10.1145/1102351.1102363).
- BURGES C. J. (2010). *From RankNet to LambdaRank to LambdaMART : An Overview*. Rapport interne MSR-TR-2010-82, Microsoft.
- CAO Z., QIN T., LIU T.-Y., TSAI M.-F. & LI H. (2007). Learning to rank : from pairwise approach to listwise approach. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, p. 129–136, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/1273496.1273513](https://doi.org/10.1145/1273496.1273513).
- CLARKE C. L., KOLLA M., CORMACK G. V., VECHTOMOVA O., ASHKAN A., BÜTTCHER S. & MACKINNON I. (2008). Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, p. 659–666, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/1390334.1390446](https://doi.org/10.1145/1390334.1390446).
- DE BORDA J. C. (1781). *Mémoire sur les élections au scrutin*, In *Histoire de l'Académie royale des sciences*, p. 657–665. Imprimerie royale de Paris.
- DING J., XIANG T., OU Z., ZUO W., ZHAO R., LIN C., ZHENG Y. & LIU B. (2022). Tell me how to survey : Literature review made simple with automatic reading path generation. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, p. 3426–3438. DOI : [10.1109/ICDE53745.2022.00322](https://doi.org/10.1109/ICDE53745.2022.00322).
- EKSTRAND M. D., KANNAN P., STEMPER J. A., BUTLER J. T., KONSTAN J. A. & RIEDL J. T. (2010). Automatically building research reading lists. In *Proceedings of the fourth ACM conference on Recommender systems*, p. 159–166, Barcelona Spain : ACM. DOI : [10.1145/1864708.1864740](https://doi.org/10.1145/1864708.1864740).
- FABBRI A., LI I., TRAIRATVORAKUL P., HE Y., TING W., TUNG R., WESTERFIELD C. & RADEV D. (2018). TutorialBank : A manually-collected corpus for prerequisite chains, survey extraction and resource recommendation. In I. GUREVYCH & Y. MIYAO, Éd., *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, p. 611–620, Melbourne, Australia : Association for Computational Linguistics. DOI : [10.18653/v1/P18-1057](https://doi.org/10.18653/v1/P18-1057).
- FESSLER J., FICARO E., CLINTHORNE N. & LANGE K. (1997). Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. *IEEE Transactions on Medical Imaging*, **16**(2), 166–175. DOI : [10.1109/42.563662](https://doi.org/10.1109/42.563662).

- FIGUEIRA P., BELEM F., ALMEIDA J. M. & GONCALVES M. A. (2019). Automatic Generation of Initial Reading Lists : Requirements and Solutions. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, p. 1–10, Champaign, IL, USA : IEEE. DOI : [10.1109/JCDL.2019.00011](https://doi.org/10.1109/JCDL.2019.00011).
- FREUND Y., IYER R., SCHAPIRE R. E. & SINGER Y. (2003). An efficient boosting algorithm for combining preferences. *J. Mach. Learn. Res.*, **4**, 933–969.
- FRIEDMAN J. H. (2001). Greedy function approximation : A gradient boosting machine. *The Annals of Statistics*, **29**(5), 1189 – 1232. DOI : [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451).
- GORDON J., AGUILAR S., SHENG E. & BURNS G. (2017). Structured Generation of Technical Reading Lists. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, p. 261–270, Copenhagen, Denmark : Association for Computational Linguistics. DOI : [10.18653/v1/W17-5029](https://doi.org/10.18653/v1/W17-5029).
- GORDON J., ZHU L., GALSTYAN A., NATARAJAN P. & BURNS G. (2016). Modeling Concept Dependencies in a Scientific Corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, p. 866–875, Berlin, Germany : Association for Computational Linguistics. DOI : [10.18653/v1/P16-1082](https://doi.org/10.18653/v1/P16-1082).
- HE Q., KIFER D., PEI J., MITRA P. & GILES C. L. (2011). Citation recommendation without author supervision. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM '11*, p. 755–764, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/1935826.1935926](https://doi.org/10.1145/1935826.1935926).
- HE Q., PEI J., KIFER D., MITRA P. & GILES L. (2010). Context-aware citation recommendation. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, p. 421–430, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/1772690.1772734](https://doi.org/10.1145/1772690.1772734).
- HO T. K. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, p. 278–282 vol.1. DOI : [10.1109/ICDAR.1995.598994](https://doi.org/10.1109/ICDAR.1995.598994).
- JARDINE J. & TEUFEL S. (2014). Topical PageRank : A model of scientific expertise for bibliographic search. In S. WINTNER, S. GOLDWATER & S. RIEZLER, Éd., *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, p. 501–510, Gothenburg, Sweden : Association for Computational Linguistics. DOI : [10.3115/v1/E14-1053](https://doi.org/10.3115/v1/E14-1053).
- JARDINE J. G. (2014). *Automatically generating reading lists*. Rapport interne UCAM-CL-TR-848, University of Cambridge, Computer Laboratory. DOI : [10.48456/tr-848](https://doi.org/10.48456/tr-848).
- KAUR J., YUSOF M., BOURSIER P. & OGIER J.-M. (2010). Automated scientific document retrieval. In *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, volume 5, p. 732–736. DOI : [10.1109/ICCAE.2010.5451344](https://doi.org/10.1109/ICCAE.2010.5451344).
- KENDALL M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**(1-2), 81–93. DOI : [10.1093/biomet/30.1-2.81](https://doi.org/10.1093/biomet/30.1-2.81).
- KLEINBERG J. M. (1999). Authoritative sources in a hyperlinked environment. *J. ACM*, **46**(5), 604–632. DOI : [10.1145/324133.324140](https://doi.org/10.1145/324133.324140).
- KUMAR R. & VASSILVITSKII S. (2010). Generalized distances between rankings. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, p. 571–580, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/1772690.1772749](https://doi.org/10.1145/1772690.1772749).
- LARSEN P. O. & VON INS M. (2010). The rate of growth in scientific publication and the decline in coverage provided by science citation index. *Scientometrics*, **84**(3), 575–603. DOI : [10.1007/s11192-010-0202-z](https://doi.org/10.1007/s11192-010-0202-z).
- LEE D. D. & SEUNG H. S. (2000). Algorithms for non-negative matrix factorization. In *Proceedings of the 13th International Conference on Neural Information Processing Systems, NIPS'00*, p. 535–541, Cambridge, MA, USA : MIT Press.

- LEMPEL R. & MORAN S. (2001). Salsa : the stochastic approach for link-structure analysis. *ACM Trans. Inf. Syst.*, **19**(2), 131–160. DOI : [10.1145/382979.383041](https://doi.org/10.1145/382979.383041).
- LEVENSHTAIN V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, **10**, 707–710.
- LI I., FABBRI A. R., TUNG R. R. & RADEV D. R. (2019). What should i learn first : Introducing lecturebank for nlp education and prerequisite chain learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**(01), 6674–6681. DOI : [10.1609/aaai.v33i01.33016674](https://doi.org/10.1609/aaai.v33i01.33016674).
- LIKERT R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, **22** **140**, 55.
- LIU H., MA W., YANG Y. & CARBONELL J. (2016). Learning concept graphs from online educational data. *J. Artif. Int. Res.*, **55**(1), 1059–1090.
- LIU T.-Y. (2009). Learning to rank for information retrieval. *Found. Trends Inf. Retr.*, **3**(3), 225–331. DOI : [10.1561/15000000016](https://doi.org/10.1561/15000000016).
- MANNING C. D., RAGHAVAN P. & SCHÜTZE H. (2008). *Introduction to information retrieval*. Cambridge University Press.
- PAGE L., BRIN S., MOTWANI R. & WINOGRAD T. (1998). *The PageRank Citation Ranking : Bringing Order to the Web*. Rapport interne, Stanford Digital Library Technologies Project.
- PAN L., LI C., LI J. & TANG J. (2017). Prerequisite relation learning for concepts in MOOCs. In R. BARZILAY & M.-Y. KAN, Éd.s., *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, p. 1447–1456, Vancouver, Canada : Association for Computational Linguistics. DOI : [10.18653/v1/P17-1133](https://doi.org/10.18653/v1/P17-1133).
- RIJSBERGEN C. J. V. (1979). *Information Retrieval*. USA : Butterworth-Heinemann, 2nd édition.
- ROBERTSON S., WALKER S., HANCOCK-BEAULIEU M. M., GATFORD M. & PAYNE A. (1996). Okapi at trec-4. In *The Fourth Text REtrieval Conference (TREC-4)*, p. 73–96 : Gaithersburg, MD : NIST.
- SEGEV A. (1987). The node-weighted steiner tree problem. *Networks*, **17**(1), 1–17. DOI : <https://doi.org/10.1002/net.3230170102>.
- SESAGIRI RAAMKUMAR A., FOO S. & PANG N. (2017). Using author-specified keywords in building an initial reading list of research papers in scientific paper retrieval and recommender systems. *Information Processing & Management*, **53**(3), 577–594. DOI : [10.1016/j.ipm.2016.12.006](https://doi.org/10.1016/j.ipm.2016.12.006).
- SIDDALL G. & ROSE H. (2014). Reading lists–time for a reality check ? an investigation into the use of reading lists as a pedagogical tool to support the development of information skills amongst foundation degree students. *Library and Information Research*, **38**(118), 52–73. DOI : [10.29173/lirg605](https://doi.org/10.29173/lirg605).
- SPEARMAN C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, **15**(1), 72–101. DOI : [10.2307/1412159](https://doi.org/10.2307/1412159).
- TANG J. & ZHANG J. (2009). A discriminative approach to topic-based citation recommendation. In T. THEERAMUNKONG, B. KIJSIRIKUL, N. CERCONE & T.-B. HO, Éd.s., *Advances in Knowledge Discovery and Data Mining*, p. 572–579, Berlin, Heidelberg : Springer Berlin Heidelberg. DOI : [10.1007/978-3-642-01307-2\\_55](https://doi.org/10.1007/978-3-642-01307-2_55).
- TANG Y. (2008). *The design and study of pedagogical paper recommendation*. Thèse de doctorat, University of Saskatchewan.
- THELWALL M. & SUD P. (2022). Scopus 1900–2020 : Growth in articles, abstracts, countries, fields, and journals. *Quantitative Science Studies*, **3**(1), 37–50. DOI : [10.1162/qss\\_a\\_00177](https://doi.org/10.1162/qss_a_00177).

THOMPSON L., MAHON C. & THOMAS L. (2004). *Reading lists - how do you eat yours ?*, In *Learning and Teaching Projects 2003-2004*, p. 57–62. University of Wolverhampton.

VARGAS S. (2014). Novelty and diversity enhancement and evaluation in recommender systems and information retrieval. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '14, p. 1281, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/2600428.2610382](https://doi.org/10.1145/2600428.2610382).

WHITE S. & SMYTH P. (2003). Algorithms for estimating relative importance in networks. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, p. 266–275, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/956750.956782](https://doi.org/10.1145/956750.956782).

XU J. & LI H. (2007). Adarank : a boosting algorithm for information retrieval. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, p. 391–398, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/1277741.1277809](https://doi.org/10.1145/1277741.1277809).