
Quels facteurs de pertinence pour la recherche de produits e-commerce ?

Laure Soulier* — Lamjed Ben Jabeur** — Paul Mousset**
— Lynda Tamine**

* Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France

laure.soulier@lip6.fr

** IRIT, Université de Toulouse, CNRS, INPT, UPS, UT1, UT2J, France, 118 Route Narbonne, Toulouse, France

{jabeur,mousset,tamine}@irit.fr

RÉSUMÉ. Un moteur de recherche e-commerce vise à fournir un accès rapide et efficace à des produits qui correspondent aux besoins et aux préférences de l'utilisateur parmi une liste de produits similaires ou étroitement liés. Nous avons participé à la campagne d'évaluation « Living Lab for Information Retrieval » qui proposait une tâche de recherche de produits évaluée par des utilisateurs réels lors de scénarios de recherche réelle sur un site de e-commerce. L'évaluation expérimentale a montré des résultats prometteurs de notre modèle. Dans ce papier, nous proposons une analyse des fichiers logs issus de notre modèle afin d'identifier des facteurs d'efficacité liés à la requête et aux produits. L'objectif de cette étude est d'ouvrir des pistes de recherche pour la formalisation de modèles de recherche de produits.

ABSTRACT. E-commerce product retrieval aims to provide a quick and efficient access to products that fit user's needs and preferences among a tail of similar or closely related products. We participated to the "Living Lab for Information Retrieval" evaluation campaign devoted to a product search task in which real users evaluated participants' retrieval models in real search scenarios on e-commerce websites. The experimental evaluation has shown encouraging results for our proposed model. In this paper, we conduct an analysis of users' feedback with respect to the clicks obtained by our model. The goal of the paper is therefore to identify the effectiveness factors underlying the user's queries and the retrieved products in order to open perspectives in the formalization of product search models.

MOTS-CLÉS : Recherche d'information, recherche de produits, facteurs d'efficacité

KEYWORDS: Information retrieval, product search, e-commerce, effectiveness factors

1. Introduction

Ces dernières années, les sites de commerce en ligne, appelés également sites de e-commerce, ont connu une croissance régulière en termes de popularité, mais également du point de vue de leurs bénéfices. Par exemple, le site leader mondial de e-commerce Amazon revendique plus de 240 millions de produits disponibles à la vente sur la version US¹. Amazon a vendu plus de 2 milliards de produits dans le monde pour l'année 2014². Ainsi, l'expansion de ce marché confronte les utilisateurs à une multitude de produits disponibles en ligne avec des prix toujours plus concurrentiels, qui renforcent la difficulté des utilisateurs à effectuer un choix. En effet, la taille importante du catalogue de produits proposés, aussi bien en termes de types et de caractéristiques, complique l'expérience d'achat des clients sur les sites de e-commerce. Pour aider les utilisateurs, la plupart des sites de e-commerce proposent des outils de recherche de produits directement intégrés dans leurs sites Web, et non plus comme des fonctionnalités optionnelles. Par conséquent, la recherche de produits est devenue un domaine de recherche à part entière, conduisant à proposer des outils de recherche adaptés afin d'aider les clients à trouver les produits qui les intéressent. Un exemple d'outil de recherche de produits est proposé par Google Shopping³ dont l'utilité a été démontrée avec 100 milliards de requêtes de recherche soumises par mois⁴.

Dans la littérature, la recherche de produits a été abordée comme une tâche de recherche d'information (RI) ayant pour objectif de relier les données de e-commerce avec le besoin en information d'un potentiel client. Alors que la RI classique repose sur des hypothèses d'appariement termes à termes entre une requête utilisateur et une collection de documents, la recherche de produits se retrouve confrontée à la difficulté de l'estimation d'une pertinence multi-dimensionnelle sur la base de produits caractérisés par un ensemble de méta-données (e.g., le titre, la description, le genre ou le prix). Dans ce contexte, des travaux ont proposé d'intégrer plusieurs caractéristiques produits de deux types. Une première catégorie de travaux, orientée « produits », se concentre principalement sur les méta-données des produits, à savoir leur catégorie et leur description, en proposant des approches de diversification de résultats de recherche (Chen *et al.*, 2011) ou de classification de produits à partir d'ontologies sémantiques (Vandic *et al.*, 2012). La deuxième catégorie de travaux repose sur un processus centré utilisateur en s'appuyant sur les préférences des utilisateurs et leurs intentions (Duan *et al.*, 2013).

Du point de vue de l'évaluation de ces modèles, une première initiative dans le cadre de la campagne « Living Labs for IR » (LL4IR) (Schuth *et al.*, 2015) à

1. <http://www.amazon.com/>

2. <http://www.ecommercebytes.com/cab/abn/y14/m07/i15/s04>

3. <http://www.google.fr/shopping>

4. <http://www.godatafeed.com/resources/google-shopping-campaigns>

CLEF 2015 (Mothe *et al.*, 2015) a vu le jour pour formaliser un cadre d'expérimentation centré utilisateur permettant à des vrais clients d'une plateforme de e-commerce⁵ de juger des produits identifiés comme pertinents par les participants de la campagne d'évaluation. Dans ce contexte, nous avons proposé un modèle probabiliste (Ben Jabeur *et al.*, 2015) qui allie à la fois l'approche orientée « produits », en se basant sur leurs descripteurs, ainsi que l'approche centrée utilisateur en prenant en compte la notion d'engagement des utilisateurs vis-à-vis d'un produit. Les résultats de la campagne d'évaluation ont souligné l'efficacité de notre approche avec des mesures d'efficacité supérieures à l'ensemble des modèles proposés par les participants ainsi que le système de référence (*baseline*).

Étant donné les résultats encourageants obtenus, nous nous intéressons dans ce papier à l'identification des facteurs d'efficacité de notre modèle pour la recherche de produits à partir des fichiers *logs* de notre modèle. Notre objectif est donc d'analyser les résultats obtenus lors de notre participation à la campagne LL4IR afin d'étudier quels facteurs, en termes de méta-données de produits, impactent sur l'efficacité de la recherche de produits.

La structure du papier est organisée comme suit. La Section 2 introduit une synthèse des travaux de recherche de produits. La section 3 présente le cadre d'expérimentation LL4IR ainsi que notre participation à cette campagne d'évaluation. La section 4 détaille notre analyse des facteurs d'efficacité pour la recherche de produits. Enfin, la section 5 conclue le papier et présente les perspectives de recherche.

2. Recherche de produits : synthèse des travaux

Alors que certains travaux se sont intéressés à la recherche de produits sous l'angle de la modélisation utilisateur pour la prise de décision (Chen, 2010; Castagnos *et al.*, 2010), d'autres travaux, plus proches de notre contribution, ont abordé la recherche de produits comme une tâche de RI qui vise exploiter les données de e-commerce pour répondre aux besoins en information exprimés par la clientèle pendant le processus d'achat en ligne.

Detlor et al. (Detlor *et al.*, 2003) ont fait une analyse comparative de deux tâches d'exploration et de recherche de produits sur les sites de e-commerce. A la différence de la tâche d'exploration dans laquelle l'utilisateur n'a pas un objectif précis pour ses achats et dont les spécifications des produits attendus sont encore vagues, la tâche recherche des produits est motivée par le fait que l'utilisateur cherche une information précise dans les spécifications d'un produit afin de prendre une décision. Outre les informations élémentaires sur le produit (par exemple le prix et la description) et les informations sur le commerçant lui-

5. <http://www.regiojatek.hu/>

même, les spécifications d'un produit sont considérées comme des informations pertinentes dans la tâche de recherche de produits.

Peu de travaux axés sur les modèles d'ordonnement se sont intéressés à la tâche de recherche de produits. Ces travaux ont principalement proposé d'intégrer plusieurs fonctionnalités que l'on peut classifier selon deux catégories.

D'une part, certains travaux se concentrent principalement sur le produit, à savoir sa catégorie et sa description. Chen et al. (Chen *et al.*, 2011) ont proposé de diversifier les résultats de recherche de produits en prenant en compte les catégories de produits et les valeurs d'attributs comme un critère de diversité. Vandic et al. (Vandic *et al.*, 2012) étudient le vocabulaire utilisé pour décrire le même produit et proposent une classification hiérarchique des produits en fonction de ces termes. En complément de cette méthode et en se basant sur une ontologie sémantique, ils proposent de reconnaître les produits similaires et de les classer selon une taxonomie de produit universelle.

D'autre part, la deuxième catégorie de travaux s'est intéressée à la problématique de recherche des produits axée sur la différence du vocabulaire utilisé sur la fiche produits et du vocabulaire utilisé par les utilisateurs pour formuler leurs requêtes de recherche. Cela inclut des requêtes de type "*PC gamer pas cher*" visant un besoin spécifique par rapport à une caractéristique technique de produit. Considérant la recherche de produits comme une tâche de recherche d'entités, Duan et al. (Duan *et al.*, 2013) proposent d'exploiter la structure attribut-valeur de la fiche de description du produit ainsi que les commentaires des utilisateurs afin de répondre aux requêtes qui relèvent des exigences techniques et qui utilisent des vocabulaires différents. Les résultats sont ainsi ordonnés en utilisant un modèle probabiliste qui estime la pertinence au niveau des préférences d'attributs.

L'objectif de ce papier est de fournir une meilleure compréhension de la tâche de recherche de produits. Contrairement à Detlor et al. (Detlor *et al.*, 2003) qui favorisent une analyse axée sur les comportements de recherche de produits, nous proposons dans ce papier une analyse orientée vers le domaine de la RI en nous intéressant aux facteurs de pertinence sous-jacents aux requêtes soumises et aux produits sélectionnés lors d'une recherche de produits sur un site de e-commerce. Pour cela, nous nous appuyons sur un modèle proposé lors de la campagne d'évaluation « Living Lab », qui allie à la fois l'approche centrée « produits » et l'approche centrée « utilisateurs », proposées dans les travaux précédents (Chen *et al.*, 2011 ; Duan *et al.*, 2013 ; Vandic *et al.*, 2012).

3. Cadre Experimental de la campagne « Living Lab »

La campagne d'évaluation "*Living Labs for Information Retrieval*" (LL4IR) (Schuth *et al.*, 2015 ; Cappellato *et al.*, 2015) propose une approche novatrice pour l'évaluation de modèles de RI. En effet, le cadre d'évaluation est carac-

térisé par l'implication d'utilisateurs dans des scénarios d'utilisation réels pour lesquels l'utilisateur soumet sa propre requête sur le site et interagit en temps réel avec les résultats retournés au moyen des *clics*. La période d'évaluation réalisée selon des scénarios de recherche réelle est appelée « période de test ». L'objectif de cette campagne est de tenir compte des *clics* générés par des utilisateurs réels afin de comparer les différents systèmes proposés par les participants et le système de production. Plus particulièrement, pour la campagne 2015, la tâche de recherche principale est une tâche de recherche de produits sur le site de commerce en ligne du leader hongrois de vente de jouets pour enfants *REGIO JÁTÉK*⁶. Dans ce papier, nous nous concentrons sur la période de test ayant eu lieu entre le 15 et 31 juillet 2015. Nous décrivons dans ce qui suit le cadre expérimental mis en œuvre lors de la campagne LL4IR 2015 en introduisant les données expérimentales, le protocole d'évaluation ainsi que notre participation avec une présentation du modèle proposé et des résultats obtenus.

3.1. *Données expérimentales*

La campagne d'évaluation LL4IR 2015 fournit une plate-forme de référence implémentée sous forme d'un service en ligne qui met à disposition des participants un ensemble de données expérimentales :

– « *participant/query* » qui regroupe 50 requêtes orientées produit. Elles sont extraites à partir des requêtes utilisateurs soumises dans le passé afin de garantir qu'elles seront vraisemblablement resoumises durant les périodes de test de la campagne d'évaluation.

– « *participant/doclist* » qui regroupe une collection de produits (appelés aussi documents). Cette collection inclut des produits disponibles, mais également ceux notés comme indisponibles qui seront potentiellement disponibles dans le futur. Le nombre moyen de produits retournés pour chaque requête est d'environ 62 produits. Ce nombre peut dépasser 500 produits pour certaines requêtes. Chaque produit est représenté par un ensemble de méta-données structurelles et sémantiques, comme les personnages associés au produit (par exemple, Barbie, Spiderman, Hello Kitty), sa marque (par exemple, Beados, LEGO, Simba) ou encore l'âge et le genre recommandé. Ces méta-données, sont présentées dans le Tableau 1.

– « *participant/feedback* » qui rassemble les jugements de pertinence mis à jour sur la plate-forme toutes les 5 minutes sur l'ensemble d'une période de test donnée. Chaque jugement de pertinence est binaire en fonction de si le produit présenté à l'utilisateur a été cliqué.

– « *participant/historicalfeedback* » qui regroupe l'historique des *clics* utilisateurs disponible pour toutes les requêtes. Cet historique est généré sur une

6. <http://www.regiojatek.hu/>

Méta-données	Description
Nom	Nom du produit
Description	Description longue du produit
Description courte	Courte description du produit
Age maximum	Âge maximum recommandé
Age minimum	Âge minimum recommandé
Date d'arrivée	Nombre de jours de stockage
Prix	Prix sans réduction
Bonus	Indique si le produit bénéficie (1) ou non (0) d'une réduction
Réduction appliquée	Montant de la réduction du produit
Pourcentage de réduction	Taux de réduction du produit
Marque	Marque
Catégorie	Catégorie du produit
Nombre de personnages	Liste des personnages connus associés au produit
Genre	Genre (0 : Neutre ; 1 : Garçon ; 2 : Fille)
Nombre de photos	Liste des photos

Tableau 1. *Description des produits*

période de plusieurs mois et a permis aux organisateurs de la campagne d'évaluation de constituer la collection de produits selon le critère qu'un produit peu cliqué est exclu de la collection.

Nous soulignons que ces différentes données expérimentales, hormis les jugements de pertinence des utilisateurs, sont statiques pour une période de test. Ces données sont mises à jour peu avant chaque période de test.

3.2. *Participants LL4IR*

Quatre organismes sont impliqués dans la tâche de recherche de produits de la campagne LL4IR. Nous présentons les différentes participations et nous détaillerons la description de notre modèle.

– «GESIS» (Schaer et Tavakolpoursaleh, 2015) ont proposé un modèle en deux étapes : (1) identifier les produits pertinents et (2) les ré-ordonner à partir de l'historique des jugements de pertinence. Cette approche de recherche a ensuite été améliorée par des techniques de filtrage sur des critères de disponibilité de produits ou de taux de réduction des prix affichés.

– «UiS» (Ghirmatsion et Balog, 2015) se sont appuyés sur des techniques de modèles de langue pour estimer la similarité produit-requête avec une intégration d'une probabilité a priori basé sur l'historique des *clics*. Trois méthodes ont été testées au travers de trois participants différents et sont présentées dans les résultats sous la forme «UiS-*».

– «Baseline» représente le système de référence proposé par les organisateurs de la campagne d'évaluation. Ce système est basé sur l'historique des *clics* (Schuth *et al.*, 2015).

– «Notre modèle» (Ben Jabeur *et al.*, 2015), notre participation, qui consiste en un modèle probabiliste utilisant les champs descriptifs textuels (e.g., titre, marque, catégorie du produit). Compte tenu de la représentation des produits sous forme des champs (attributs et valeurs) et inspirés par les travaux de Craswell *et al.* (Craswell *et al.*, 2005) et Dakka *et al.* (Dakka *et al.*, 2012), nous proposons de représenter un produit p par deux ensembles des champs dont (1) l'ensemble de champs descriptifs textuels \mathcal{D}_p qui décrivent le produit p et (2) l'ensemble des champs de structuration c_p et permettent d'organiser les produits selon des catégories. La pertinence $P(p|q)$ d'un produit p par rapport à la requête q est traduite par la probabilité $P(c_p, \mathcal{D}_p|q)$ (Equation 1). Selon la transformation des probabilités de Bayes (Equation 2) et compte tenu de l'indépendance assumée entre la catégorie d'un produit et sa description (Equation 3), la pertinence d'un produit est estimée par l'équation suivante :

$$P(p|q) = P(c_p, \mathcal{D}_p|q) \quad [1]$$

$$= P(c_p|q) \cdot P(\mathcal{D}_p|c_p, q) \quad [2]$$

$$\propto P(c_p|q) \cdot P(\mathcal{D}_p|q) \quad [3]$$

Avec $P(c_p|q)$ représentant respectivement la pertinence de la catégorie c_p du produit p en se basant sur une distribution de la catégorie c_p sur l'ensemble des produits. $P(\mathcal{D}_p|q)$ exprime la pertinence thématique des champs descriptifs \mathcal{D}_p du produit p par rapport à la requête q au travers du modèle *BM25F*. Ce modèle calcule la similarité entre le produit p et la requête q tout en attribuant l'importance variée à chacun des champs. Le cadre d'évaluation imposé par le LL4IR ne permettant aux participants de soumettre l'ordonnancement d'un seul modèle, nous nous sommes appuyés sur les expérimentations menées dans les précédents travaux afin de fixer les poids ((Zaragoza *et al.*, 2004) pour les poids des descripteurs et (Craswell *et al.*, 2005; Zaragoza *et al.*, 2004) pour les poids des paramètres du *BM25F*). Les produits pertinents sont ensuite réordonnés selon une notion d'engagement des utilisateurs traduite autour d'une notion de la popularité (Lehmann *et al.*, 2012). Cette dernière est estimée par le nombre de marquages « j'aime » et de partages générés sur la plate-forme sociale Facebook⁷.

3.3. Protocole d'évaluation

Afin d'évaluer les modèles de recherche des produits via la plate-forme LL4IR, les participants procèdent aux étapes suivantes :

7. <http://www.facebook.com>

- 1) Recueillir l'ensemble des requêtes depuis la ressource requêtes "*participant/query*"
- 2) Pour chaque requête, extraire la liste des produits candidats, disponible via la ressource doclist "*participant/doclist*".
- 3) Pour chaque ID de produit dans la liste, récupérer le contenu du produit par la ressource API "*participant/doc*".
- 4) Pour chaque requête, soumettre la liste des produits ordonnée selon un modèle de recherche de produits via la ressource "*participant/run*".

En conformité avec l'approche centrée utilisateur proposée par la campagne LL4IR 2015, les ordonnancements des produits proposés par l'ensemble des participants sont mélangés avec l'ordonnement par défaut du système de production du site de e-commerce. Lorsqu'un utilisateur soumet une requête qui appartient à l'ensemble des requêtes pré-sélectionnées sur la plate-forme, l'utilisateur reçoit un ensemble de produits dont la moitié provient du système de production du site web et l'autre moitié des ordonnancements fourni par les participants à la campagne d'évaluation.

Dans l'objectif d'évaluer chacune des approches, les organisateurs de la campagne LL4IR 2015 ont proposé cinq métriques. Ces dernières sont estimées sur la base de l'ensemble des requêtes soumises :

- Le nombre de « victoires », noté $\#Wins$, qui exprime le nombre de fois où le système du participant a reçu plus de *clics* que le système de production.
- Le nombre de « pertes », noté $\#Losses$, qui exprime le nombre de fois où le système du participant a reçu moins de *clics* que le système de production.
- Le nombre d'« égalités », noté $\#Ties$, qui exprime le nombre de fois où le système du participant a reçu autant de *clics* que le système de production.
- Le nombre d'« impressions », noté $\#Impressions$, qui exprime le nombre de fois où un ordonnancement d'un participant a été mélangé à celui du système de production. Plus particulièrement, le nombre d'impressions est estimé comme suit :

$$\#Impressions = \#Wins + \#Losses + \#Ties \quad [4]$$

– Le « ratio de victoires », noté $Outcome$, est défini comme le quotient de « victoires » sur la sommes de « victoires » et « pertes », comme indiqué dans l'équation 5. Un ratio supérieur à 0.5 souligne la capacité du système à fournir plus de produits pertinents que de produits non pertinents, en se basant sur l'hypothèse que les *clics* sont des indicateurs de pertinence (Joachims, 2002).

$$Outcome = \frac{\#Wins}{\#Wins + \#Losses} \quad [5]$$

3.4. Résultats obtenus pour la campagne LL4IR

Le Tableau 2 présente les résultats obtenus par les différents participants. En excluant le premier participant n’ayant pas fourni des ordonnancements sur toutes les requêtes, les résultats soulignent l’efficacité de notre modèle. En effet, nous obtenons la valeur de « ratio de victoires » la plus élevée (0.489) qui dépasse les valeurs des métriques obtenues pour l’ensemble des participants ainsi que la *Baseline* fournie par les organisateurs de la campagne d’évaluation (Schuth *et al.*, 2015). Ces modèles reposent essentiellement sur les historiques des *clics* ainsi que les descripteurs des produits; ces derniers étant utilisés également dans notre modèle. Alors que l’historique des *clics* aurait pu être un facteur significatif de l’engagement des utilisateurs, les augmentations obtenues pour notre modèle (comprises entre +8.49% et +41.37%) mettent en évidence la pertinence de notre estimation de l’engagement utilisateur au sein d’une tâche de recherche de produits. De plus, notre modèle a obtenu des valeurs de « ratio de victoires » très proches de 0.5, suggérant que notre modèle est très proche de celui de système de production déployé sur le site de e-commerce. Les valeurs de « ratio de victoires » obtenues par les participants demeurent cependant inférieures à 0,5; soulignant la difficulté de formaliser des modèles de recherche de produits qui est un domaine de recherche relativement jeune. Cependant, les résultats encourageants obtenus par notre modèle nous poussent à effectuer des analyses complémentaires afin de mieux comprendre les facteurs de pertinence pour la recherche de produits. Nous aborderons cet aspect dans la prochaine section.

	Outcome	#Wins	#Losses	#Ties	#Impressions
Peter Dekker *	0.5429	19	16	76	111
Notre modèle	0.4890	89	93	533	715
UiS-Mira	0.4507	64	78	527	669
Baseline	0.4430	66	83	498	647
GESIS	0.4134	74	105	513	692
UiS-Jern	0.3702	67	114	511	692
UiS-UiS	0.3459	55	104	521	680

Tableau 2. Résultats LL4IR sur la période 15-31 juillet 2015. * : participant n’ayant pas fourni des ordonnancements sur toutes les requêtes.

4. Analyse des facteurs de pertinence pour la recherche de produits

Dans cette section, nous nous intéressons particulièrement à l’analyse des facteurs de pertinence afin d’ouvrir des pistes de recherche pour la formalisation de modèles de recherche de produits. Ne disposant pas des fichiers log du système de production pour cette période d’évaluation, nous nous appuyons sur les résultats obtenus par notre modèle en faisant l’hypothèse qu’il constitue une bonne approximation des préférences des utilisateurs dans la mesure où il a

obtenu des résultats sensiblement équivalents à ceux du système de production. Dans ce contexte, nous nous intéresserons particulièrement aux deux questions de recherche suivantes :

– **RQ1** : Quelles sont les caractéristiques des requêtes utilisateurs lors d’une recherche de produits ? Quel est l’impact de ces caractéristiques sur l’efficacité de la recherche ?

– **RQ2** : Quelles sont les caractéristiques des produits retournés par notre modèle ? Quelles caractéristiques des produits impactent l’efficacité de la recherche ?

4.1. Facteurs de pertinence liés à la requête

Nous proposons de nous intéresser ici à la première question de recherche (RQ1) axée sur l’analyse des caractéristiques des requêtes ainsi que les facteurs sous-jacents qui peuvent impacter la pertinence de chacune des requêtes. Dans un premier temps, nous nous intéressons à caractériser les requêtes générées dans le cas de recherches de produits et formulées par des utilisateurs réels. Le Tableau 3 présente les caractéristiques des 50 requêtes évaluées lors de la campagne LL4IR. Ce sont des requêtes généralement courtes, voire très courtes (1 mot) qui font référence à des jeux et objets ou également des produits dérivés de marques (par exemple « Monster », « Lego » ou « Playmobil »). Peu de requêtes évoquent des personnages tels que « Angry Bird » et « Cars ».

	Caractéristique	Nombre de requêtes
Long.	1 mot	45
	2 mots	5
Classe	Object, dispositif et jeux	31
	Marque, franchise, ligne de produit et modèle de produit	10
	Thème, catégorie et concept	5
	Personnage fictif, film, servie TV et jeux vidéos	3
	Animal	1
	Autres	1
Pers.	Avec personnage	7
	Sans personnage	53
Marq.	Avec marque	10
	Sans marque	40

Tableau 3. *Analyse descriptive des caractéristiques des requêtes*

Nous nous intéressons maintenant à l’analyse des relations entre les caractéristiques et l’efficacité de ces requêtes. Dans un premier temps, nous présentons au travers de la Figure 1 l’efficacité de chaque requête par rapport à la mesure « ratio de victoires » fournie à l’issue de la campagne d’évaluation LL4IR. Nous soulignons que parmi les 50 requêtes, nous avons obtenu un « ratio de victoires » supérieur ou égal à 0.5 pour 26 requêtes. Cependant, nous observons

également 17 requêtes avec une mesure d'efficacité égale à 0. Nous remarquons ainsi qu'il existe trois classes de requêtes identifiées selon le critère d'efficacité permettant de classifier les requêtes suivant si elles obtiennent des résultats supérieurs ou inférieurs au système de production ou encore des requêtes en situation d'échec. Afin de construire des classes de requêtes de façon formelle, nous procédons à une classification *k-means* en trois classes (*3-means*), permettant d'identifier les requêtes suivant leur échec, succès ou si elles ont obtenu des résultats acceptables en termes d'efficacité. Le Tableau 4 résume les statistiques d'efficacité sous-jacentes à chaque classe. La classification est bien conforme à notre intuition, fixant des seuils entre les classes à 0,2 et 0,6 pour accéder respectivement aux classes « Acceptable » et « Succès ».

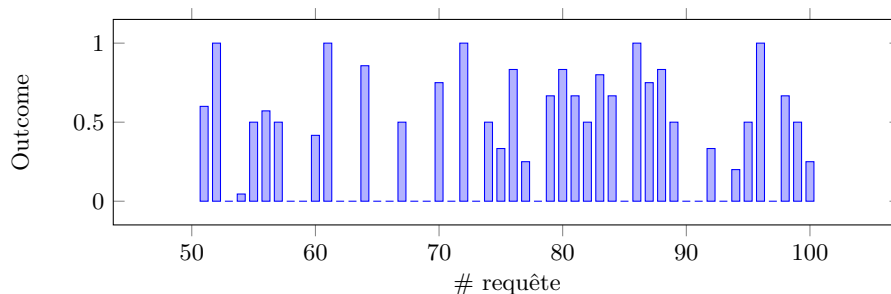


Figure 1. Analyse des requêtes selon la mesure d'efficacité « ratio de victoires »

	Nombre de requêtes	Min	Moyenne	Max	Écart-type
Échec	19	0	0,0129	0,200	0,046
Acceptable	15	0,250	0,450	0,600	0,109
Succès	16	0,667	0,833	1	0,133

Tableau 4. Analyse de l'efficacité selon chaque classe de requêtes

Nous nous sommes ensuite intéressés à définir le modèle statistique qui permet d'expliquer quelles caractéristiques ont un impact significatif sur l'efficacité. Pour cela, nous avons considéré la variable issue de la classification *3-means* comme variable à expliquer et les caractéristiques présentées précédemment dans le Tableau 3 comme variables explicatives. Les résultats montrent qu'aucune des caractéristiques liées à la requête impacte l'efficacité obtenue. Ces résultats peuvent être expliqués par le fait que les requêtes ne sont pas assez nombreuses dans chaque catégorie des caractéristiques (voir Tableau 3), ne permettant pas d'avoir des différences significatives entre les classes.

4.2. Facteurs de pertinence liés au produit

La synthèse des travaux du domaine ainsi que les modèles expérimentés lors de la campagne d'évaluation illustre l'importance des caractéristiques des

produits (plutôt que de la requête) dans la prise en compte du score de pertinence. Ainsi, dans ce qui suit, nous nous intéressons à étudier les relations potentielles entre les caractéristiques des produits et l’efficacité. Plutôt que de déduire l’efficacité à partir de la mesure « ratio de victoires » estimée au niveau de la requête et dénotant plutôt une préférence de l’utilisateur au système de référence, nous souhaitons exploiter les jugements de pertinence des utilisateurs pour chaque produit retourné par notre modèle. Cette analyse permet d’identifier une relation entre les caractéristiques des produits et les préférences des utilisateurs vis-à-vis de chaque produit afin d’approximer au mieux les spécificités de la recherche produit. Pour cette raison, plutôt que d’utiliser l’historique des jugements de pertinence qui sont fortement corrélés à la temporalité (« effet de mode » d’un produit) et à la disponibilité des produits, nous utilisons une mesure fournie par les organisateurs du « Living Lab » correspondant à la probabilité d’un produit d’être cliqué. Cette probabilité est estimée durant la période de test par le ratio entre le nombre de *clicks* d’un produit et le nombre de fois où le produit a été présenté à l’utilisateur.

Le Tableau 5 présente les caractéristiques qualitatives des produits. Les variables numériques sont présentées dans le Tableau 6. Les produits généralement retournés par notre modèle sont des produits assexués qui ne bénéficient pas spécifiquement de réduction. Le site étant un site de vente de jouets, les produits recommandés sont généralement préférés pour des enfants entre 3 et 6 ans, présentés avec 2 photos et rarement reliés à des personnages connus. Le prix moyen de ces produits est de 348.7 forints hongrois (équivalent à 1.10€) avec un taux moyen de réduction à 13.14%.

	Caractéristique	Nombre de produits
Genre	Neutre	2397
	Fille	368
	Garçon	267
Bonus	Avec réduction	292
	Sans réduction	2740

Tableau 5. *Analyse descriptive des caractéristiques qualitatives des produits*

Caractéristique	Min	Moyenne	Max
Age minium	0	3.387	10
Age maximum	0	6	10
Date d’arrivée	20	932.1	2098
Prix	1	348.7	712
Réduction appliquée	0	309.3	3000
Pourcentage de réduction	1	13.14	252
Nombre de photos	0	1.968	50
Nombre de personnages	0	0.12	2

Tableau 6. *Analyse descriptive des caractéristiques numériques des produits*

De façon analogue à l'analyse des requêtes, notre objectif est de construire le modèle statistique qui permettrait d'expliquer quels sont les produits que les utilisateurs ont tendance à acheter sur les sites de e-commerce. Étant donné les deux types de variables (qualitatives et quantitatives), nous avons effectué une analyse de covariance où nous avons construit le modèle itérativement en enlevant à chaque itération le facteur avec la *p-value* la plus importante jusqu'à n'inclure que des caractéristiques avec des *p-values* inférieures ou égales à 0.5. Les résultats sont présentés dans le Tableau 7.

caractéristiques	coefficient	p-value
Nombre de photos	2.367e-3	8.01e-07 ***
Date d'arrivée	5.242e-07	0.001 **
Pourcentage de réduction	-4.349e-02	0.009 **
Réduction appliquée	1.744e-02	0.008 **

Tableau 7. *Modèle descriptif des préférences utilisateurs lors de la recherche de produits sur des sites de e-commerce*

L'analyse de covariance révèle que trois classes de caractéristiques de produits impactent les préférences des utilisateurs : la présence de photo(s), l'ancienneté d'un produit et également les informations relatives à son prix. En effet, la présence de photos semble susciter l'intérêt des utilisateurs, comportement traduit par un coefficient positif dans l'analyse de covariance. Très simplement, ce résultat est expliqué par le fait que la présence de photos apporte plus de crédibilité au produit et permet aux utilisateurs d'avoir une image du produit qu'ils souhaitent acheter. En effet, la présence de photos est un facteur déclenchant d'achat bien connu des stratégies de marketing avec certaines études mettant en avant le fait que 67% de consommateurs considèrent que les images des produits sont extrêmement importantes lors de leurs achats en ligne⁸. La date d'arrivée du produit semble également impacter positivement le comportement des utilisateurs. Les résultats semblent suggérer que les utilisateurs apprécient les produits avec un peu d'ancienneté dans la mesure où ils attendent un petit peu de temps avant de les acheter afin d'avoir l'avis des autres utilisateurs qui auront acheté le produit dès les premiers jours de sa mise à disposition.

Une autre conclusion qu'il est intéressant de souligner réside dans le fait que les caractéristiques liées au prix et à la réduction du prix d'un produit impactent également les préférences des utilisateurs. Dans le même esprit, bien que l'analyse de covariance révèle que la présence d'une réduction génère une appétence vers le produit plus forte, la valeur de la réduction, quant à elle, semble être négativement corrélée avec les préférences des utilisateurs. D'une façon intéressante, le prix semble positivement corrélé ainsi que le taux de réduction négativement corrélé aux préférences des utilisateurs permettent de déduire que la recherche de produits est particulièrement orientée vers des

8. <http://blog.lemonstand.com/7-ways-optimize-product-page-conversions/>

produits spécifiques (par exemple « Playmobil ») et que les utilisateurs ne recherchent donc pas de produits similaires moins onéreux.

Il nous semble également important de souligner que certaines caractéristiques n'ont pas été identifiées comme significatives en termes d'impact sur les préférences des utilisateurs. Il s'agit principalement des caractéristiques de genre, d'âge recommandé et des personnages connus reliés au produit.

4.3. Discussion

L'analyse des facteurs de pertinence repose sur des métriques qui compare le modèle évalué avec le système de production. Ces mesures comportent certaines limites dans la mesure où elles ne sont pas basées sur des mesures formelles et ne prennent pas en compte le nombre de clics global des utilisateurs. Malgré cela, notre analyse a permis de mettre en évidence les conclusions suivantes :

– *Des requêtes non discriminantes pour la recherche de produits.* Les analyses descriptives semblent montrer dans un premier temps que les requêtes formulées par les utilisateurs n'incluent pas spécifiquement des caractéristiques « produit » (telles que la marque ou le personnage). Cette intuition a été démontrée par l'analyse de variance qui a approuvé qu'il est difficile de mesurer l'inefficacité de la recherche, voire qu'il n'existe pas de relation entre les caractéristiques des requêtes et l'efficacité de la recherche. Cependant, nous sommes conscients qu'il aurait fallu étayer ces résultats par une analyse comparative des requêtes réalisées dans le cas d'une recherche produit et une recherche documentaire évaluée selon un même cadre d'évaluation ; ce qui n'a pas été mis en place dans le cas de la campagne d'évaluation « Living Lab ».

– *Des préférences utilisateurs orientées sur une gamme de produits plutôt que la recherche de produits génériques.* Les résultats mettant en avant l'impact du prix et du taux de réduction sur la probabilité de pertinence d'un produit ont mis en évidence le besoin des utilisateurs d'identifier des produits correspondant exactement à leur requête plutôt que de favoriser une recherche de produits similaires et moins onéreux. De façon similaire, l'absence de relation entre la pertinence du produit et les caractéristiques de genre et d'âge suggèrent que les utilisateurs recherchent des produits bien spécifiques et que favoriser un genre ou un âge en particulier ne permettra pas de provoquer une déviance dans le comportement de l'utilisateur. En effet, une requête orientée produit est souvent reliée à un produit ou type de produit en particulier (par exemple « doll »), ainsi l'utilisateur a déjà une idée du genre, de l'âge recommandé ou des personnages reliés au produit. Cependant, cette explication est difficile à démontrer dans la mesure où les requêtes ne sont rattachées à aucune donnée de ce type, empêchant la mise en correspondance entre les préférences des utilisateurs et les produits jugés comme pertinents.

– *La présentation du produit est prépondérante.* La pertinence des produits étant impactée par la présence de photos dans la description du produit signifie la prépondérance de l'aspect visuel sous-jacent à la présentation du pro-

duit. Par conséquent, il serait intéressant de prendre en compte la présentation des caractéristiques des produits comme facteur explicatif de la pertinence du produit. En effet, les réductions sont généralement affichées sous forme d'un pictogramme qui signifie à l'utilisateur que le prix est soldé. Ceci peut expliquer les coefficients opposés obtenus pour les facteurs liés à l'absence (ou la présence) d'une réduction et le taux de réduction. Un aspect intéressant aurait été d'analyser l'impact de ce pictogramme sur l'appréciation de la présence de réduction, en effectuant la même évaluation sur un site ne favorisant pas cette mise en avant de la réduction.

Les conclusions précédentes nous permettent d'identifier le besoin de formaliser des modèles de recherche de produits plus ciblés et moins orientés similarité. En effet, dans ce cas de requêtes orientées produits, l'objectif d'un modèle d'ordonnancement n'est donc pas, comme dans le cas d'une tâche de recherche d'information, d'identifier des produits fortement similaires (de marques différentes par exemples) qui gravitent autour de la description, mais plutôt de fournir à l'utilisateur l'ensemble des produits répondant spécifiquement au besoin. Par exemple, dans le cas de la requête « Playmobil », l'utilisateur attendra plus particulièrement tous les jouets de la gamme « Playmobil » et non des jouets génériques, tels que des bonhommes en plastiques du type soldat, animaux, etc... De même, un modèle retournant des produits favorisant la diversité des genres ou des âges recommandés ne sera probablement pas pertinent pour ce type de besoin, ce qui est différent en recherche d'information classique où l'utilisateur a besoin de documents diversifiés pour assurer une couverture du besoin en information.

5. Conclusion

Dans cet article, nous nous sommes intéressés à comprendre les facteurs de pertinence pour la recherche de produits. Cette étude fait suite à notre participation à la campagne « Living Lab » qui propose un protocole expérimental novateur basé sur des utilisateurs en situation de recherche réelle. Lors de cette campagne d'évaluation, nous avons proposé un modèle probabiliste basé sur la recherche de produits reposant sur les caractéristiques descriptives des produits et l'engagement des utilisateurs. Les résultats ont montré l'efficacité de notre modèle vis-à-vis des modèles proposés par l'ensemble des participants.

Forts de ces résultats, nous nous sommes appuyés sur les fichiers *logs* de notre modèle afin d'identifier les facteurs de pertinence liés à la requête et aux produits. Les résultats n'ont montré aucune relation entre l'efficacité d'une requête et ses caractéristiques, suggérant une formulation du besoin utilisateur similaire à un besoin en information classique. De par la nature de la tâche orientée vers un acte d'achat de produits, nous avons pu révéler que les facteurs de pertinence sont fortement dépendant des caractéristiques des produits, plutôt que de la requête. Parmi les facteurs significatifs, nous avons identifié des facteurs liés à la description visuelle du produit (présence de photo), son an-

cienneté sur le marché et également son prix. Ce dernier révèle que l'utilisateur ne cherche pas nécessairement un produit similaire à la requête (et moins onéreux) comme dans le cas d'une recherche classique, mais plus particulièrement une gamme de produits. Ainsi, nous espérons que ces conclusions permettront de formaliser des modèles de recherche de produits plus ajustés aux spécificités de cette tâche.

6. Bibliographie

- Ben Jabeur L., Soulier L., Tamine L., « IRIT at CLEF 2015 : A product search model for head queries », *CLEF 2015 Workshop*, 2015.
- Cappellato L., Ferro N., Jones G., San Juan E. e. ., « CLEF 2015 Labs and Workshops, Notebook Papers », *CEUR Workshop*, 2015.
- Castagnos S., Jones N., Pu P., « Eye-tracking product recommenders' usage », *Proceedings of the fourth ACM conference on Recommender systems*, ACM, 2010.
- Chen L., « Social influence of product popularity on consumer decisions : usability study of Flickr camera finder », *IUI*, p. 297-300, 2010.
- Chen X., Wang H., Sun X., Pan J., Yu Y., « Diversifying Product Search Results », *SIGIR*, ACM, p. 1093-1094, 2011.
- Craswell N., Robertson S., Zaragoza H., Taylor M., « Relevance Weighting for Query Independent Evidence », *SIGIR*, ACM, p. 416-423, 2005.
- Dakka W., Gravano L., Ipeirotis P., « Answering General Time-Sensitive Queries », *IEEE ToKDE*, vol. 24, p. 220-235, 2012.
- Detlor B., Sproule S., Gupta C., « Pre-Purchase Online Information Seeking : Search versus Browse. », *J. Electron. Commerce Res.*, vol. 4, n° 2, p. 72-84, 2003.
- Duan H., Zhai C., Cheng J., Gattani A., « Supporting Keyword Search in Product Database : A Probabilistic Approach », *VLDB*, vol. 6, n° 14, p. 1786-1797, 2013.
- Ghirmatsion A. B., Balog K., « Probabilistic field mapping for product search », *CLEF 2015 Workshop*, 2015.
- Joachims T., « Optimizing Search Engines Using Clickthrough Data », *KDD*, p. 133-142, 2002.
- Lehmann J., Lalmas M., Yom-Tov E., Dupret G., « Models of User Engagement », *UMAP*, Springer, p. 164-175, 2012.
- Mothe J., Savoy J., Kamps J., Pinel-Sauvagnat K., Jones G., San Juan E., Cappellato L., Ferro N. e., « Experimental IR meets Multilinguality, Multimodality, and Interaction », *CLEF*, LNCS, vol. 9283, Springer, 2015.
- Schaer P., Tavakolpoursaleh N., « Historical Clicks for Product Search : GESIS at CLEF LL4IR 2015 », *CLEF 2015 Workshop*, 2015.
- Schuth A., Balog K., Kelly L., « Overview of the Living Labs for Information Retrieval Evaluation (LL4IR) CLEF Lab 2015 », *CLEF*, Springer, 2015.
- Vandic D., van Dam J.-W., Frasinca F., « Faceted product search powered by the Semantic Web », *Decision Support Systems*, vol. 53, n° 3, p. 425 - 437, 2012.
- Zaragoza H., Craswell N., Taylor M. J., Saria S., Robertson S. E., « Microsoft Cambridge at TREC 13 : Web and Hard Tracks. », Citeseer, 2004.