
Prise en compte des préférences des utilisateurs pour l'estimation de la pertinence multidimensionnelle d'un document

Bilel Moulahi^{*,} — Lynda Tamine^{*} — Sadok Ben Yahia^{**,***}**

** Université Paul Sabatier, Institut de Recherche en Informatique de Toulouse, France
bilel.moulahi@irit.fr, lynda.tamine@irit.fr*

*** Université de Tunis El Manar, Faculté des Sciences de Tunis, LIPAH, 2092, Tunisie
sadok.benyahia@fst.rnu.tn*

**** Institut Mines-Télécom, Télécom SudParis, UMR CNRS Samovar, 91011 Evry Cedex, France*

RÉSUMÉ. Dans ce papier, nous proposons une nouvelle approche d'agrégation personnalisée pour l'estimation de la pertinence multidimensionnelle. L'approche est basée sur un opérateur d'agrégation mathématique qui utilise une mesure floue permettant la quantification de l'importance estimée des critères pour chaque utilisateur ainsi que leur degré d'interactivité ou d'interdépendance. Nous évaluons l'opérateur d'agrégation proposé en utilisant la collection de test standard fournie avec par la tâche "Contextual Suggestion" de TREC 2013. Les résultats expérimentaux obtenus montrent l'impact de la personnalisation sur les performances de recherche.

ABSTRACT. In this paper, we propose a novel personalized aggregation approach to the multidimensional relevance aggregation. The approach is based on a mathematical aggregation operator relying on a fuzzy measure that allow to quantify the importance degree of each relevance dimension for every user as well as the interaction existing between the criteria. The evaluation of our approach is performed on the test collection of TREC 2013 Contextual Suggestion track. Experimental results show the impact of the personalisation of search results.

MOTS-CLÉS : Personnalisation, préférences, pertinence, Choquet personnalisé, capacité.

KEYWORDS: Personalization, preferences, relevance, personalized Choquet, capacity.

1. Introduction

De nombreux travaux en recherche d'information (RI) ont mis en exergue à la fois l'importance et la complexité du concept "*pertinence*" (Borlund, 2003, Saracevic, 2007, Taylor *et al.*, 2007). Son importance est liée au fait que la notion sous-jacente est le fondement des modèles d'ordonnement de documents en réponse à une requête, qui est la finalité même d'un système de RI (Baeza-Yates *et al.*, 1999). Sa complexité est, quant à elle, subordonnée à deux propriétés. La première concerne la multiplicité de ses dimensions, vues comme des ensembles de critères, qui peuvent être de surcroît, interdépendantes ; même si de nombreux travaux du domaine se sont focalisés sur la dimension thématique seule, force est de constater que de nombreux autres travaux ont prouvé empiriquement l'impact conjoint de plusieurs dimensions sur l'estimation de la pertinence finale, comme la tâche et la situation de recherche (Borlund, 2003, Saracevic, 2007, Taylor *et al.*, 2007). Considérons à titre d'exemple, une tâche de recherche de *tweets* ; des analyses expérimentales ont montré que la pertinence d'un *tweet* en réponse à une requête, est impactée principalement par la conjonction de trois dimensions qui sont le sujet et la fraîcheur du *tweet* et l'autorité du *tweeter* qui l'a émis (Nagmoti *et al.*, 2010). La seconde propriété concerne la subjectivité qui entoure ces dimensions ; en effet, la plupart d'entre elles ne sont pas basées sur des estimations objectives puisqu'elles sont fortement liées à la perception personnelle des utilisateurs impliqués dans la tâche de RI ; on cite à titre d'exemple les centres d'intérêt, l'expertise et les préférences des utilisateurs. La problématique scientifique est alors de définir des opérateurs capables d'agrèger des scores de pertinence partiels (relatifs à chaque dimension) en tenant compte de leur interdépendance éventuelle. Cette problématique a été abordée dans diverses applications de RI comme la RI personnalisée (Sieg *et al.*, 2007, Tamine *et al.*, 2006, Daoud *et al.*, 2010), la RI mobile (Göker *et al.*, 2008), la RI sociale (Nagmoti *et al.*, 2010) et la RI géographique (Mata *et al.*, 2011). Cependant, ces travaux applicatifs ont généralement utilisé des opérateurs de calcul de moyenne pondérée ou de combinaison linéaire qui se basent sur l'hypothèse non réaliste d'additivité ou d'indépendance des dimensions. D'autres travaux fondamentaux récents, se sont intéressés en revanche à la définition d'opérateurs d'agrégation, indépendamment du cadre applicatif, qui permettent de traiter peu ou prou le biais de l'interactivité (da Costa Pereira *et al.*, 2012, Gerani *et al.*, 2012, Eickhoff *et al.*, 2013). Toutefois, ces opérateurs ne permettent pas de tenir compte de la propriété de subjectivité qui peut se décliner à travers les différences entre les utilisateurs quant à l'importance accordée à chaque dimension de pertinence. Notre contribution, présentée dans ce papier, répond à cet objectif. Plus précisément, nous proposons un opérateur flou d'agrégation basé sur l'intégrale de Choquet (Choquet, 1953, Grabisch, 1995), capable d'agrèger des scores de pertinence personnalisés, puisque pondérés par l'importance orientée-utilisateur de chaque dimension.

La suite du papier est organisée comme suit : la section 2 présente un aperçu des travaux du domaine et situe notre contribution dans ce contexte. La section 3 détaille les principes de l'opérateur d'agrégation ainsi que l'algorithme d'apprentissage des mesures d'importance. La section 4 décrit le cadre expérimental puis les résultats

de l'application de l'opérateur proposé dans une tâche TREC dédiée à une tâche de RI personnalisée en l'occurrence "TREC¹ Contextual Suggestion" (Dean-Hall *et al.*, 2013).

1.1. Synthèse des travaux

Le concept de pertinence est incontestablement au centre d'une activité de recherche d'information comme en témoignent les nombreux travaux qui en ont fait l'objet d'étude (Saracevic, 1976, Borlund, 2003, Saracevic, 2007). L'un des résultats phares qui ressort de ces travaux est que la pertinence est estimée en globalité selon un ensemble de dimensions qui s'apparentent à des familles de critères ; parmi ces différentes dimensions, on cite les plus reconnues dont : la pertinence thématique (contenu et méta-contenu), la pertinence situationnelle (temps et géolocalisation) et la pertinence cognitive (expertise, centres d'intérêts). Un autre résultat important est l'interdépendance de ces dimensions pour inférer la pertinence globale d'un document (Nagmoti *et al.*, 2010, Saracevic, 2007). En clair, un utilisateur juge de la pertinence d'un document en tenant compte conjointement de l'ensemble des critères de pertinence ; à titre d'exemple, un document est d'autant plus pertinent du point de vue du contenu que l'expertise de l'utilisateur est en lien avec ce contenu. Historiquement, la dimension thématique est particulièrement considérée dans le domaine. La prise en compte de la propriété de multiplicité des dimensions de pertinence a particulièrement émergé dans des cadres applicatifs de la RI comme :

- la RI mobile (Göker *et al.*, 2008) : un document est d'autant plus pertinent pour une requête qu'il en est proche thématiquement et qu'il comporte des liens vers des lieux géographiquement proches de l'utilisateur qui est en situation de mobilité ;
- la RI sociale (Nagmoti *et al.*, 2010) : un document (ou ressource sociale) est d'autant plus pertinent pour une requête qu'il en est proche thématiquement, qu'il émane d'un acteur socialement important et qu'il est recommandé par un ami ;
- la RI personnalisée (Sieg *et al.*, 2007, Daoud *et al.*, 2007, Daoud *et al.*, 2010) : un document est d'autant plus pertinent pour une requête qu'il en est proche thématiquement et qu'il est en adéquation avec les centres d'intérêts de l'utilisateur ;
- RI géographique (Daoud *et al.*, 2013) : un document est d'autant plus pertinent pour une requête qu'il en est proche thématiquement et qu'il comporte des liens vers des lieux géographiquement proches des lieux cités dans la requête ;

La plupart de ces travaux exploitent des opérateurs classiques de produit, de moyenne pondérée et de combinaison linéaire. D'autres travaux (Palacio *et al.*, 2010) exploitent des opérateurs de combinaison inspirés de la fusion des données. Cependant ces opérateurs répondent à la problématique de l'agrégation en se basant sur l'hypothèse d'additivité ou d'indépendance des dimensions de pertinence. D'autres travaux récents ont particulièrement examiné le principe d'agrégation de dimensions interac-

1. <http://trec.nist.gov>

tives indépendamment du cadre applicatif (da Costa Pereira *et al.*, 2012, Gerani *et al.*, 2012, Eickhoff *et al.*, 2013). Celia *et al.* (da Costa Pereira *et al.*, 2012) ont proposé un opérateur d'agrégation multidimensionnelle mettant en jeu quatre (4) critères de pertinence : contenu, couverture, adéquation et fiabilité en définissant deux opérateurs d'agrégation prioritaire en l'occurrence, "And" et "Scoring". Ces opérateurs modélisent un ordre de priorité entre les critères de pertinence sur la base d'un mode de calcul de poids associés qui favorise la satisfaction du critère d'ordre supérieur ; les travaux présentés dans (Bouidghaghen *et al.*, 2011) ont montré l'efficacité de ces opérateurs dans un cadre de RI mobile. Gerani *et al.* (Gerani *et al.*, 2012) ont proposé un opérateur qui ne nécessite pas la satisfaction de la condition de comparabilité des scores partiels de pertinence. Ils utilisent à cet effet un algorithme de transformation de scores basé sur l'algorithme *Alternating Conditional Expectation* et le modèle *Box-Cox*. Plus récemment, Eickhoff *et al.* (Eickhoff *et al.*, 2013) ont proposé une approche statistique basé sur la méthode *Copulas* qui traite spécifiquement la complexité des dépendances des critères de pertinence. Eickhoff *et al.* ont montré que la méthode *Copulas* permet de modéliser des relations de dépendances complexes entre les différentes dimensions de pertinence. Leur approche a été évaluée dans trois tâches de RI à savoir, la recherche d'opinions dans les blogs, la recherche personnalisée dans les folksonomies et la recherche web adaptées aux enfants.

1.2. Aperçu de la contribution et positionnement

La cadre général de nos travaux concerne l'agrégation de dimensions de pertinence, qu'elles soient interdépendantes ou indépendantes, dans une tâche de RI. Plus spécifiquement, nous présentons une approche d'agrégation personnalisée des scores de pertinence basée sur l'usage d'une mesure floue, appelée capacité, sous-jacente à l'opérateur de Choquet (Choquet, 1953). Cette mesure est à la base de la quantification de l'importance estimée de chaque dimension pour chaque utilisateur ainsi que leur degré d'interactivité ou d'interdépendance ; elle est estimée selon un algorithme d'apprentissage, qui infère les mesures optimales en utilisant une vérité de terrain évaluable à l'aide de la métrique précision de la recherche. Nous évaluons l'opérateur d'agrégation proposé en utilisant la collection de test standard TREC Contextual Suggestion (Dean-Hall *et al.*, 2013) et montrons l'impact de la prise en compte des dépendances entre critères ainsi que leur personnalisation sur les performances de recherche. Comparativement aux travaux antérieurs proches (da Costa Pereira *et al.*, 2012, Gerani *et al.*, 2012, Eickhoff *et al.*, 2013) ainsi qu'à notre précédente contribution (Moulahi *et al.*, 2013), le travail présenté dans ce papier s'en distingue selon les principaux points clés suivants :

- 1) une agrégation pondérée par les préférences des utilisateurs quant à chacune des dimensions agrégées, contrairement aux travaux de l'état de l'art présentés dans (da Costa Pereira *et al.*, 2012, Gerani *et al.*, 2012, Eickhoff *et al.*, 2013) ainsi que dans notre précédente contribution (Moulahi *et al.*, 2013) ; ces travaux proposent de déployer des opérateurs produisant des scores de pertinence dépendant seulement des

dimensions de pertinence agrégées, indépendamment des utilisateurs ;

2) un nouvel algorithme d'apprentissages des mesures d'importance des critères, comparativement à l'algorithme présenté dans (Moulahi *et al.*, 2013) ;

3) une nouvelle évaluation expérimentale tant dans l'objectif que dans la méthodologie, qui montre, comparativement à celle menée dans (Moulahi *et al.*, 2013), à la fois l'intérêt de l'agrégation et de la personnalisation des préférences des utilisateurs sur les performances de recherche.

2. Agrégation personnalisée de la pertinence multidimensionnelle

2.1. Formalisation de l'opérateur d'agrégation

Nous introduisons le problème d'agrégation de pertinence multidimensionnelle comme étant un problème de prise de décision multicritères où les critères considérés sont les dimensions de pertinence. En effet, le défi majeur dans le problème d'agrégation est : (1) l'estimation de l'importance des critères : identifier les critères devant avoir un poids d'importance plus élevé que d'autres ; (2) l'agrégation : combiner efficacement les critères de pertinence en tenant compte des dépendances pouvant exister entre eux.

Soient $\mathcal{D} = \{d_1, d_2, \dots, d_M\}$ un ensemble de documents, $\mathcal{C} = \{c_1, c_2, \dots, c_N\}$ l'ensemble des critères de pertinence et q une requête donnée. La tâche de combinaison des critères notée $RSV_{c_i}^u(q, d_j)$, d'un document $d_j \in \mathcal{D}$, obtenu suivant chaque critère de pertinence $c_i \in \mathcal{C}$, est appelé *agrégation*. La fonction \mathcal{F} qui calcule le score de pertinence personnalisé du document d_j en réponse à la requête q , d'un utilisateur u , a la forme suivante :

$$\mathcal{F} : \begin{cases} \mathbb{R}^N \longrightarrow \mathbb{R} \\ (RSV_{c_1}^u(q, d_j) \times \dots \times RSV_{c_N}^u(q, d_j)) \longrightarrow \mathcal{F}(RSV_{c_1}^u(q, d_j), \dots, RSV_{c_N}^u(q, d_j)) \end{cases}$$

Où $RSV_{c_i}^u(q, d_j)$ est le score de pertinence de d_j suivant le critère c_i , étant donné l'utilisateur u .

Dans ce qui suit, nous allons nous baser sur l'intégrale de Choquet comme un opérateur d'agrégation de pertinence multidimensionnelle. Cette fonction mathématique est construite à l'aide d'une mesure floue (ou *capacité*) μ , définie comme suit.

Définition 1 Soit $I_{\mathcal{C}}$ l'ensemble de tous les sous ensembles de critère de \mathcal{C} . Une mesure floue est une fonction monotone normalisée μ de $I_{\mathcal{C}}$ à $[0 \dots 1]$ tels que : $\forall I_{C_1}, I_{C_2} \in I_{\mathcal{C}}$, si $(I_{C_1} \subseteq I_{C_2})$ alors $\mu(I_{C_1}) \leq \mu(I_{C_2})$, avec $\mu(I_{\emptyset}) = 0$ et $\mu(I_{\mathcal{C}}) = 1$.

Pour simplifier la notation, $\mu(I_{C_i})$ sera dénotée par μ_{C_i} . La valeur de μ_{C_1} peut être interprétée par le degré d'importance de l'interaction entre les critères inclus dans le sous ensemble C_1 . La fonction d'agrégation de pertinence personnalisée basée sur l'intégrale de Choquet est définie comme suit :

Définition 2 $RSV_{\mathcal{C}}^u(q, d_j)$ est le score de pertinence personnalisé de d_j pour l'utilisateur u suivant l'ensemble des critères de pertinence $\mathcal{C} = \{c_1, c_2, \dots, c_N\}$ défini comme : $RSV_{\mathcal{C}}^u(q, d_j) = Ch_{\mu}(RSV_{c_1}^u(q, d_j), \dots, RSV_{c_N}^u(q, d_j))$
 $= \sum_{i=1}^N \mu_{\{c_i, \dots, c_N\}}^u \cdot (rsv_{(i)j}^u - rsv_{(i-1)j}^u)$

Où Ch_{μ} la fonction d'agrégation de Choquet, $rsv_{(i)j}^u$ est le $i^{\text{ème}}$ élément de la permutation $RSV(q, d_j)$ sur le critère c_i , tel que $(0 \leq rsv_{(1)j}^u \leq \dots \leq rsv_{(N)j}^u)$, $\mu_{\{c_i, \dots, c_N\}}^u$ est le degré d'importance de l'ensemble des critères $\{c_i, \dots, c_N\}$ pour l'utilisateur u .

De cette manière, nous sommes capables d'ajuster les paramètres du modèle d'ordonnement automatiquement pour chaque utilisateur, rendant ainsi les résultats dépendants de ses préférences sur les critères considérés. Notons que si μ est une mesure additive, l'intégrale de Choquet correspond à la moyenne pondérée. Sinon, elle demande moins de 2^N mesures de capacité dans le cas où la mesure est k -additive, i.e., $\mu_A = 0$ pour tous les sous ensembles de critères $A \subseteq \mathcal{C}$ avec $|A| > k$. D'un point de vue théorique, l'intégrale de Choquet dispose d'un nombre de propriétés qui semblent être pertinentes pour un domaine tel que la RI ; étant donné qu'elle est construite à partir du concept de mesure floue, elle permet la modélisation des relations d'interaction flexibles en considérant des relations de dépendance complexes entre les critères (Grabisch *et al.*, 2000).

Pour faciliter la tâche d'interprétation du modèle résultat de l'intégrale de Choquet, nous allons exploiter deux paramètres appelés, "indice d'importance" et "indice d'interaction" (Grabisch *et al.*, 2000). L'indice d'importance, appelé également indice de Shapley, permet d'estimer la contribution moyenne qu'un critère (c_i) apporte à toutes les autres combinaisons de critères possibles. L'indice d'interaction permet de donner des informations sur le phénomène d'interaction pouvant exister entre un ensemble de critères. Pour des détails sur le calcul de ces deux indices, le lecteur peut se référer au papier original (Grabisch *et al.*, 2000).

2.2. Apprentissage des préférences des utilisateurs

L'objectif de la phase d'apprentissage est d'optimiser les mesures floues selon une mesure objective de RI (e.g. $P@X$) en identifiant les valeurs de capacité permettant de personnaliser les résultats de recherche d'un utilisateur en particulier, tout en considérant ses préférences individuelles sur les critères de pertinence.

Nous proposons dans ce qui suit un algorithme générique permettant d'apprendre ces capacités indépendamment du nombre de critères de pertinence, et de la tâche de RI considérée. Étant donné un utilisateur, les données d'apprentissage pour identifier les mesures floues de l'intégrale de Choquet comprennent un ensemble de requêtes d'apprentissage, et pour chaque requête, un ensemble ordonné de documents représentés par des vecteurs contenant des scores partiels selon chaque critère ; chaque

Notation	Description
Q_{app}^u	L'ensemble des requêtes utilisées pour apprendre les valeurs de capacités de l'utilisateur u
N	Nombre de critères de pertinence
\mathcal{D}	La collection de documents
K	Nombre de documents utilisés pour l'apprentissage pour chaque requête
$\gamma^{i,r}$	Liste ordonnée de documents en réponse à la requête q_r suivant la combinaison de capacité $\mu^{(i)}$. Soit $P@X(\gamma^{r,i})$ la $P@X$ de $\gamma^{r,i}$ et $AVP@X(\gamma^i)$ soit sa moyenne de $P@X$ sur toutes les requêtes $\in Q_{app}$ suivant $\mu^{(i)}$
I_{C_r}	Tous les sous ensembles de critères possibles de C_r
\mathcal{S}_μ	Ensemble de combinaisons de capacité expérimentées. Chaque combinaison $\mu^{(i)} \in \mathcal{S}_\mu$ contient les valeurs de capacités de tous les ensembles et sous ensemble de critères

Tableau 1 – Synthèse des notations utilisées avec l'algorithme 1.

document est annoté avec une étiquette (*e.g.*, pertinent ou non pertinent). La méthodologie adoptée est détaillée dans l'algorithme 1. Le tableau 1 décrit les notations utilisées dans cet algorithme. Ce dernier comprend deux étapes principales :

– *Initialisation des valeurs initiales des combinaisons de capacités.* Une combinaison de capacités $\mu^{(\cdot)}$ désigne l'ensemble des valeurs de capacités associées à chaque critère et à chaque sous-ensemble de critères. Par exemple, dans le cas de trois critères de pertinence, une combinaison de capacités comprend $(\{\mu_{c_1}; \mu_{c_2}; \mu_{c_3}; \mu_{c_1,c_2}; \mu_{c_1,c_3}; \mu_{c_2,c_3}\})$. Afin de paramétrer ces valeurs, nous utilisons une mesure de RI telle que la $P@X$ sur les requêtes d'apprentissage Q_{app}^u . Le paramétrage est concevable étant donné que le nombre de critères de pertinence est généralement petit (Saracevic, 2007). Cependant, lorsque le nombre de critères est supérieur ou égale à 4, nous pouvons éviter la complexité du paramétrage en se basant sur la famille des capacités 2-additive (Grabisch *et al.*, 2000) nécessitant moins de coefficients à définir.

– *Optimisation des valeurs de capacités.* En partant d'une combinaison de capacités $\mu^{(*)}$ obtenue dans l'étape précédente, on extrait les K premiers documents retournés en réponse à chaque requête $q \in Q_{app}^u$. Les scores de ces documents (D_{learn}^u) sont interpolés pour placer les documents non pertinents à la fin de l'ordonnement. Après avoir obtenu les scores de pertinence globaux désirés $RSV_c^{int}(q, d_j)$ pour chaque document $d_j \in D_{app}^u$, et étant donné que nous disposons des étiquettes $RSV_{c_i}^u(q, d_j)$, nous procédons à l'application de la méthode des moindres carrés pour l'identification des valeurs de capacités des critères et des sous-ensembles de critères considérés.

Algorithm 1 Apprentissage des mesures floues

Entrées: Q_{learn}^u, N, K .

Sortie: Combinaison de capacité optimale $\mu^{(**)}$.

Étape 1 : Initialisation des valeurs de capacités

$m \leftarrow (1 - N) \times N$;

1. **Pour** $i = 1$ à m *{Identification des combinaisons de capacités}* **Faire**
2. $\mu^{(i)} = \left(\bigcup_{j:1..N} \{\mu_{c_j}\} \right) \cup \left(\bigcup_{Cr \in \mathcal{C}, |Cr| > 1} \{\mu_{I_{Cr}}\} \right)$; $\mu_{I_{Cr}} = \sum_{c_i \in Cr, |c_i|=1} \mu_{c_i}$
3. **Fin Pour**
4. **Si** $N \geq 4$ *{Supposer la 2-additivité}* **Alors**
5. **Pour** chaque $I_{Cr} \in \mu^{(i)}$ tel que $|Cr| > 2$ **Faire**
6. $\mu_{I_{Cr}} = 0$
7. **Fin Pour**
8. **Fin Si**
9. $\mathcal{S}_\mu = \bigcup_{i:1..m} \{\mu^{(i)}\}$
10. **Pour** chaque $\mu^{(i)} \in \mathcal{S}_\mu$ *{paramétrage des capacités}* **Faire**
11. Calculer $AVP@X(\gamma^i)$
12. **Fin Pour**
13. $Cmax = \underset{1..|\mathcal{S}_\mu|}{\text{Argmax}} (AVP@X(\gamma^i))$; $\mu^{(*)} = \mu^{(Cmax)}$

Étape 2 : Optimiser les valeurs de capacités

14. $D_{app}^u = \emptyset$
 15. **Pour** $r = 1$ à $|Q_{app}^u|$ *{Interpoler les scores globaux}* **Faire**
 16. $D_{app}^u = D_{app}^u \cup \gamma^{*,r}$
 17. **Pour** $j = 1$ à K **Faire**
 18. $RSV_C^{int}(q_r, d_j) = \underset{1..d'_j \in \gamma^{*,r}, d'_j > c d_j}{\text{Max}} (RSV_C^u(q_r, d'_j))$; $\gamma^{*,r} = \gamma^{*,r} \setminus \{d_j\}$
 19. **Fin Pour**
 20. **Fin Pour**
{Optimisation basée sur la méthode des moindres carrées}
 21. **Répéter**
 $\mathcal{F}_{LS}(\mu) = \sum_{d_j \in D_{learn}^u} [Ch_\mu(RSV_{c_1}^u(d_j), \dots, RSV_{c_N}^u(d_j)) - RSV_C^{int}(d_j)]^2$
 22. **Jusqu'à** convergence
 23. **Retourner** le résultat $\mu^{(**)}$
-

3. Cadre expérimental

Notre évaluation expérimentale est basée sur la collection de test standard fournie par la tâche “Contextual Suggestion” de TREC² 2013 (Dean-Hall *et al.*, 2013). Cette tâche a pour objectif d’évaluer les techniques de recherche répondant à des besoins en information, qui sont fortement tributaires du contexte et des centres d’intérêts des uti-

lisateurs. Étant donné un utilisateur, cette tâche a pour objectif de chercher les places d'attractions pouvant l'intéresser suivant deux critères de pertinence dépendants : (1) les centres d'intérêt de l'utilisateur, *i.e.*, ses préférences personnelles sur un historique de recherche de places ; (2) sa localisation géographique.

3.1. Données expérimentales

La collection de test présente les caractéristiques suivantes :

– **Utilisateurs** : le nombre total d'utilisateurs est égal à 635. Chaque utilisateur est représenté par un profil reflétant ses préférences sur des lieux d'une liste de 50 exemples de suggestions. Un exemple de suggestion est un lieu d'attraction qui est susceptible d'intéresser l'utilisateur. Chaque exemple est représenté par le titre du lieu, une brève description et une URL du site web correspondant. Les préférences des utilisateurs sont données sur une échelle de 5 points et sont attribuées aux descriptions et aux URLs des exemples de suggestions. Les préférences positives (*resp.*, négatives) sont celles ayant un degré de pertinence égal à 3 ou à 4 (*resp.*, 0 ou 1) selon la description du site et la correspondance par rapport à l'URL.

– **Contextes et requêtes** : le nombre de contextes fournis est égal à 50 ; chaque contexte correspond à une position géographique dans une ville donnée. La position géographique est décrite par une longitude et une latitude. Étant donnée une paire d'utilisateurs et un contexte représentant la requête, l'objectif principal de la tâche est de fournir une liste de 50 suggestions triée par ordre de pertinence selon les critères centres d'intérêt de l'utilisateur et géolocalisation.

– **Collection de documents** : pour chercher des suggestions de lieux à partir du web, nous avons exploité l'API Google Place³. Comme pour la plupart des groupes participant à la tâche "Contextual Suggestion" (Dean-Hall *et al.*, 2013), nous commençons par interroger l'API Google Place avec les requêtes appropriées en se basant sur la localisation géographique des lieux. Étant donné que l'API Google Place renvoie jusqu'à 60 suggestions par requête, nous avons effectué une nouvelle recherche avec des paramètres différents tels que les types de lieux qui sont pertinents par rapport à la tâche (*e.g.*, restaurant, pizzeria, musée, etc.). Nous avons collecté, en moyenne, environ 157 suggestions par requête et 3925 suggestions au total. Pour obtenir les scores des documents collectés selon le critère de géolocalisation, nous avons calculé la distance entre les lieux collectés et le contexte. Les scores des documents selon le critère centres d'intérêts est calculé en se basant sur le cosinus de similarité entre la description des suggestions et le profil de l'utilisateur. Les profils des utilisateurs sont représentés par des vecteurs de termes construits à partir de leurs préférences personnelles sur les exemples de suggestions. La description des lieux est construite à partir des "snippets" des résultats renvoyés par le moteur de recherche Google⁴ lorsque l'URL du lieu est soumise sous forme d'une requête.

3. <https://developers.google.com/places>

4. <https://www.google.com>

– **Jugements de pertinence** : les jugements de pertinence de cette tâche sont effectués par les utilisateurs et mandatés par TREC à la fois (Dean-Hall *et al.*, 2013). Chaque utilisateur représenté par un profil, juge les lieux qui lui sont suggérés de la même façon que les exemples de suggestions. Ainsi, l'utilisateur affecte un jugement de 0 – 4 à chaque titre/description et à chaque URL, tandis que les assesseurs de TREC jugent les suggestions uniquement en termes de correspondance au critère géo-localisation avec une évaluation de (2, 1 et 0). Une suggestion est considérée comme pertinente si elle a un degré de pertinence égal à 3 ou 4 selon le critère centre d'intérêts (profil) et une évaluation égale à 1 ou 2 selon le critère géolocalisation. Dans ce qui suit, ces jugements de pertinence constituent notre réalité de terrain utilisée pour l'apprentissage et le test.

3.2. Protocole d'évaluation

Nous avons adopté une méthodologie entièrement automatisée basée sur une validation croisée afin d'identifier les valeurs de capacité des utilisateurs et tester les performances du modèle d'agrégation. À cette fin, nous avons procédé à une partition aléatoire de l'ensemble des 50 contextes en deux ensembles de même taille, noté Q_{app}^u et Q_{test}^u utilisés respectivement pour l'apprentissage et le test. En outre, pour éviter le problème de surapprentissage, l'ensemble des contextes est divisé aléatoirement dans un second tour en deux ensembles différents d'apprentissage et de test.

L'objectif principal de la phase d'apprentissage est d'apprendre les capacités $(\mu_{\{centre_interet\}}^u, \mu_{\{localisation\}}^u)$ qui correspondent à l'importance des critères de pertinence. Nous commençons d'abord par une mesure floue initiale donnant le même poids d'importance pour les deux critères de pertinence. Ensuite, nous calculons la mesure de précision $P@5$ de tous les contextes de l'ensemble d'apprentissage Q_{app}^u . En utilisant la vérité de terrain fournie avec la tâche "Contextual Suggestion" de TREC 2013, et en se basant sur l'algorithme 1, nous identifions pour chaque utilisateur ses préférences personnelles sur les deux critères : centres d'intérêts et localisation géographique. Enfin, pour tester l'efficacité de notre approche, nous nous sommes appuyés sur l'ensemble de contextes restants Q_{test}^u et nous avons utilisé la mesure officielle de la tâche $P@5$ pour le calcul de performances. Cette mesure de précision est équivalente à la proportion des suggestions de lieux pertinents retournés parmi les 5 premiers.

4. Résultats expérimentaux

4.1. Analyse de l'importance des critères de pertinence

Notre premier objectif consiste à analyser les valeurs de capacité issues de l'algorithme 1, représentant le degré d'importance des critères de pertinence pour les utilisateurs $(\mu_{\{centre_interet\}}^u, \mu_{\{geolocalisation\}}^u)$. A cet effet, nous commençons par analyser l'importance intrinsèque de chaque critère indépendamment des autres cri-

tères. La figure 1 montre la variation des valeurs de capacité pour chaque utilisateur selon les deux critères de pertinence sur l'ensemble Q_{app}^u d'apprentissage. L'axe des abscisses représente l'ensemble des utilisateurs (35-669) et l'axe des ordonnées représente les valeurs de capacité correspondantes selon les critères centres d'intérêt (Ci) et géolocalisation (Geo).

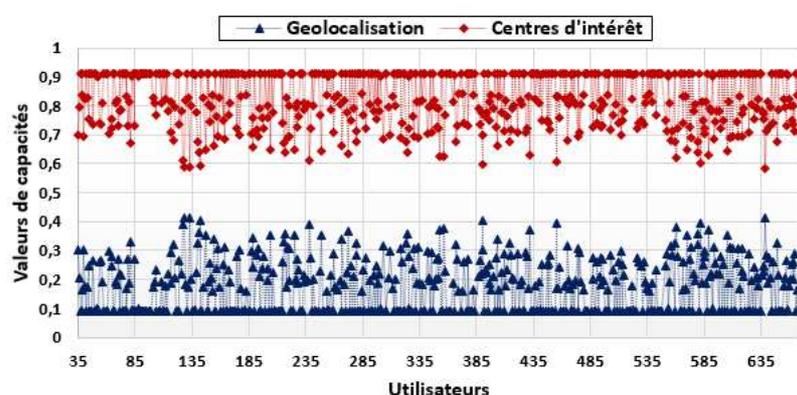


Figure 1 – Valeurs de capacités des utilisateurs de la tâche “Contextual Suggestion” de TREC 2013 suivant les deux critères de pertinence centres d'intérêt et géolocalisation.

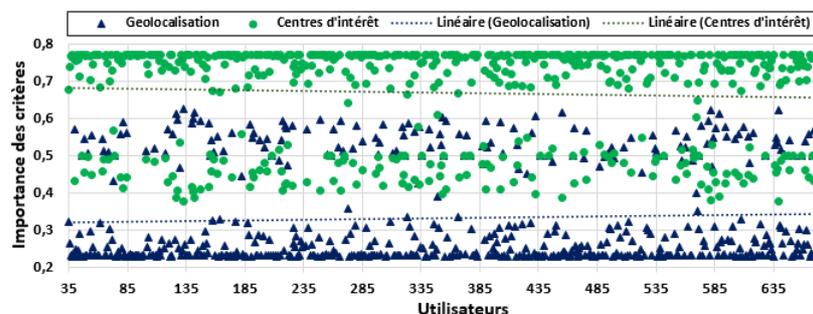


Figure 2 – Importance des critères centres d'intérêt (Ci) et géolocalisation pour les utilisateurs de la tâche “Contextual Suggestion” de TREC 2013.

En se référant à la figure 1, nous constatons que le critère Ci se voit accorder une capacité plus importante que le critère Geo . Par exemple, l'utilisateur 285 a une valeur de capacité de l'ordre de 0,23 pour le premier critère alors qu'il a une mesure de l'ordre de 0,76 pour le critère Geo . Ceci est prévisible étant donné que les utilisateurs de cette tâche s'intéressent généralement aux lieux qui correspondent à leurs préférences personnelles, même si elles ne sont pas géographiquement pertinentes. Cependant, la figure 1 montre que la distribution des valeurs de capacité est loin d'être la même pour tous les utilisateurs et met en exergue des valeurs qui vont de 0,09 à 0,414 pour le critère Geo et d'autres qui vont de 0,585 à 0,909 pour le critère Ci .

Pour mieux comprendre ce constat, nous traçons sur la figure 2, les valeurs des indices d'importance reflétant, pour chaque utilisateur, le degré de préférence globale selon les deux critères de pertinence C_i et Geo . A la différence de la figure 1, la figure 2 met en évidence l'importance moyenne de chaque critère de pertinence quand il est associé à l'autre critère. On peut observer sur la figure 2 que les préférences des utilisateurs sur les deux critères sont totalement différentes. Le lissage des valeurs d'importance obtenues selon ces critères donne deux courbes linéaires avec des valeurs tout à fait constantes et différentes, corroborant ainsi les résultats obtenus sur la figure 1. Le critère "centre d'intérêt" est encore pondéré par une importance relativement élevée pour la plupart des utilisateurs. Néanmoins, on peut également remarquer au milieu de la figure (valeurs comprises entre 0, 4 et 0, 7) que certains utilisateurs ont une préférence élevée sur le critère géolocalisation et inversement.

Dans une seconde étape, nous analysons à travers la figure 3, la dépendance entre les critères pour chaque utilisateur par le biais de l'indice d'interaction (Grabisch, 1995). Plus les valeurs de cet indice sont proches de 1 (*resp.*, -1) plus les deux critères sont dépendants et l'interaction est positive (*resp.*, négative). Si la valeur de l'indice d'interaction est égale 0, les deux critères sont considérés comme indépendants et par conséquent, il n'existe aucune interaction entre ces derniers. On peut constater que les valeurs obtenues sur tous les utilisateurs sont toutes positives et varient entre 0,28 et 0,99. La valeur moyenne est de l'ordre de 0,56 ce qui implique une interaction positive entre les deux critères de pertinence considérés lorsqu'ils sont combinés ensemble.

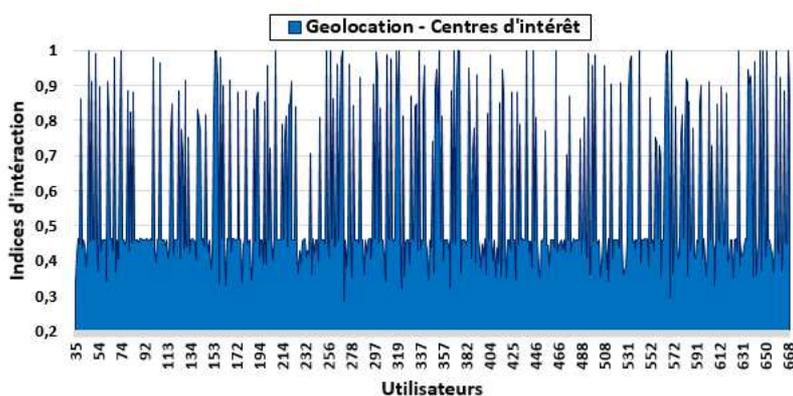


Figure 3 – Indices d'interaction entre les critères de pertinence centres d'intérêt et géolocalisation pour chaque utilisateur.

4.2. Analyse des performances de recherche

Notre second objectif est d'évaluer les performances de notre approche en termes : (i) d'agrégation de pertinence multidimensionnelle ; et (ii) de personnalisation des

préférences des utilisateurs sur les critères de pertinence. Pour ce faire, nous comparons les résultats obtenus sur l'ensemble de contextes de test Q_{test}^u aux méthodes d'agrégation de référence (*baseline*) : la moyenne arithmétique pondérée (MAP) largement utilisée dans la plupart des approches impliquant la combinaison des scores de pertinence et les deux opérateurs d'agrégation prioritaires SCORING et AND, précédemment utilisés pour l'agrégation de pertinence dans un cadre de RI personnalisée. Il convient de préciser que nous avons effectué une série d'expérimentations avec une validation croisée pour identifier les meilleurs scénarios de priorisation devant être utilisés avec les deux opérateurs SCORING et AND sur le même ensemble d'apprentissage utilisé pour trouver les valeurs de capacité de Choquet. Comme pour les résultats obtenus dans la phase d'analyse des indices d'importance, nous avons également constaté que le meilleur scénario est celui donnant une priorité au critère "centres d'intérêt" des utilisateurs. Cependant, les opérateurs d'agrégation ne sont pas en mesure de quantifier le degré d'importance des critères comme c'est le cas pour l'intégrale de Choquet.

Afin de montrer l'efficacité de l'approche de personnalisation, nous comparons notre opérateur d'agrégation personnalisé Choquet, notée CHOPER *versus* l'opérateur d'agrégation Choquet classique non personnalisé. Les capacités utilisées avec l'opérateur de Choquet classique sont obtenus en appliquant l'algorithme 1 une seule fois (et non pas pour chaque utilisateur), donnant ainsi en sortie des valeurs d'importance sur les critères indépendamment des préférences individuelles de chaque utilisateur. Ceci donne lieu à une valeur de 0,86 pour le critère centre d'intérêt et une valeur de l'ordre de 0,14 pour le critère géolocalisation. Les mesures de précision obtenues sont moyennées sur toutes les séries de tests et pour l'ensemble des requêtes de test.

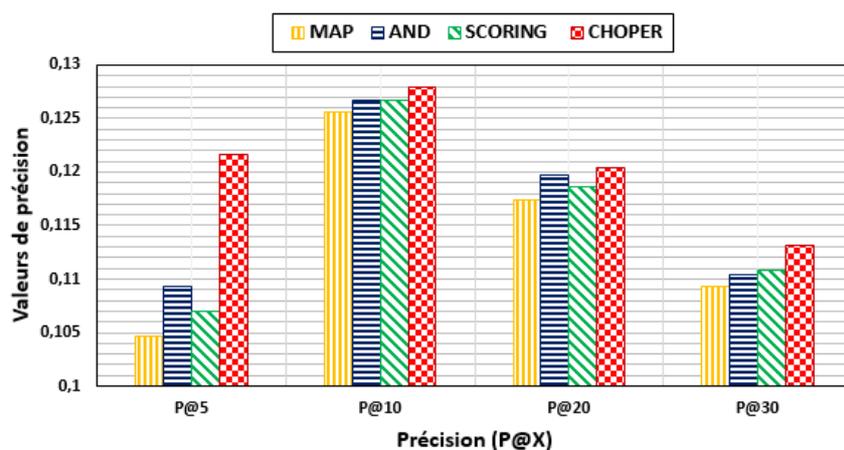


Figure 4 – Efficacité de notre approche d'agrégation de pertinence dans la tâche "Contextual Suggestion" de TREC 2013 en comparaison avec les méthodes de référence.

La figure 4 présente les résultats obtenus par notre approche CHOPER, en comparaison avec les méthodes de référence. La figure 4 montre que les performances de l'opérateur CHOPER sont significativement plus élevées que toutes les autres méthodes suivant la mesure officielle $P@5$, mais également suivant les autres mesures. Pour tester l'importance des améliorations obtenues par notre approche, nous avons effectué un t -test, et nous avons trouvé que toutes ces améliorations sont statistiquement importantes avec des p -valeurs < 0.01 pour toutes les fonctions d'agrégations testées.

La meilleure amélioration obtenue par notre approche suivant $P@5$ est marquée avec la méthode WAM (13.98%). En comparaison avec la meilleure méthode de référence (*i.e.*, AND), les améliorations sont significatives mais moins importantes (10.11%) en termes de $P@5$. Ces résultats sont probablement dus au fait que l'opérateur d'agrégation prioritaire AND est principalement basé sur l'opérateur MIN, ceci pourrait pénaliser les lieux pertinents selon le critère le moins important à savoir, le critère géolocalisation. Vu que la plupart des utilisateurs ont une préférence moins importante selon ce critère, la pénalisation de ce dernier permet d'améliorer les performances de recherche. La différence obtenue dans la performance, en faveur de CHOPER, s'explique par la prise en compte des différents niveaux de préférence suivant les deux critères de pertinence ainsi que la prise en compte de l'interaction qui existe entre ces derniers.

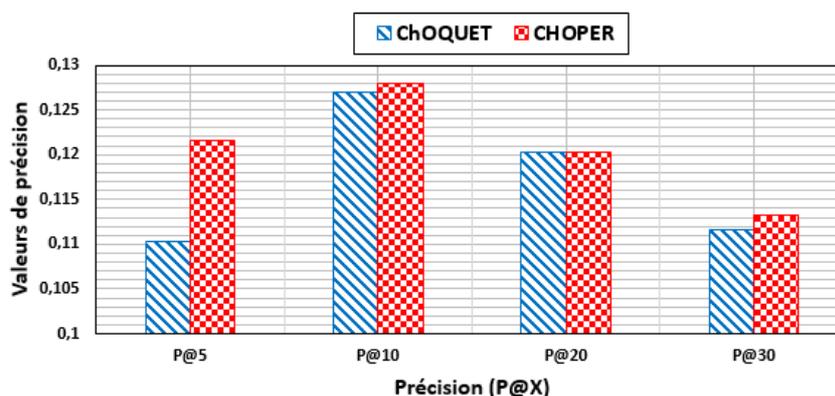


Figure 5 – Efficacité de notre approche en terme de personnalisation en comparaison avec l'opérateur d'agrégation de Choquet classique.

En termes de personnalisation, la figure 5 présente les résultats obtenus en termes de précisions ($P@5$, $P@10$, $P@20$ et $P@30$) entre l'opérateur classique Choquet et sa version personnalisée CHOPER. Ces résultats montrent que le dernier est plus performant sur toutes les mesures de précision. La meilleure amélioration est de l'ordre de 9,29% en termes de $P@5$. Ces résultats confirment ceux obtenus dans la phase d'identification des capacités (Cf. section 4.1) où nous avons montré que les degrés d'importance des critères dépend des préférences de l'utilisateur et ne sont pas les

mêmes pour tous. La prise en compte des poids d'importance appropriés pour chaque critère et chaque utilisateur permet de donner ainsi des résultats à la fois pertinents et adaptés aux préférences personnelles des utilisateurs.

5. Conclusion et perspectives

Dans ce papier, nous avons présenté une nouvelle approche pour l'agrégation de pertinence multidimensionnelle en tenant compte des préférences des utilisateurs. Notre approche repose sur une méthode d'agrégation floue permettant de pondérer les préférences des utilisateurs à chacun des critères agrégés. En se basant sur les indices d'importance et d'interaction, notre modèle permet de mesurer et donc d'interpréter les poids d'importance associés avec chaque critère de pertinence. L'évaluation de notre approche dans une tâche de recherche de lieux d'attraction et sur la collection de test fournie par la tâche "Contextual Suggestion" de TREC 2013, montre l'efficacité de notre approche dans l'agrégation multicritères et l'effet positif de la personnalisation des préférences des utilisateurs sur les résultats obtenus. En perspective, nous envisageons d'étendre l'approche de personnalisation proposée vers des groupes d'utilisateurs plutôt que des utilisateurs individuels. Ceci permettrait d'apprendre les préférences à partir des utilisateurs des classes similaires, permettant ainsi de pallier au problème d'insuffisance des exemples d'apprentissage.

6. Bibliographie

- Baeza-Yates R. A., Ribeiro-Neto B. A., *Modern Information Retrieval*, ACM Press / Addison-Wesley, 1999.
- Borlund P., « The concept of relevance in IR », *Journal of the American Society for Information Science and Technology*, vol. 54, n° 10, p. 913-925, 2003.
- Bouidghaghen O., Tamine L., Pasi G., Cabanac G., Boughanem M., da Costa Pereira C., « Prioritized Aggregation of Multiple Context Dimensions in Mobile IR », *In Proceedings of the 7th Asia conference on Information Retrieval Technology*, vol. 7097 of AIRS'11, Springer, Berlin, Heidelberg, p. 169-180, 2011.
- Choquet G., « Theory of capacities », *Annales de l'Institut Fourier*, vol. 5, p. 131-295, 1953.
- da Costa Pereira C., Dragoni M., Pasi G., « Multidimensional relevance : Prioritized aggregation in a personalized Information Retrieval setting », *Information Processing and Management*, vol. 48, n° 2, p. 340-357, 2012.
- Daoud M., Huang J. X., « Modeling geographic, temporal, and proximity contexts for improving geotemporal search », *Journal of the American Society for Information Science*, vol. 64, n° 1, p. 190-212, 2013.
- Daoud M., Tamine L., Boughanem M., « A Personalized Graph-Based Document Ranking Model Using a Semantic User Profile », *In Proceedings of the 18th international conference on User Modeling, Adaptation, and Personalization*, UMAP'10, Berlin, Heidelberg, p. 171-182, 2010.

- Daoud M., Tamine L., Boughanem M., Chebaro B., « Learning Implicit User Interests Using Ontology and Search History for Personalization », *Proceedings of the 2007 International Conference on Web Information Systems Engineering, WISE'07*, Springer-Verlag, Berlin, Heidelberg, p. 325-336, 2007.
- Dean-Hall A., Clarke C., Kamps J., Thomas P., Simone N., Voorhes E., « Overview of the trec 2013 contextual suggestion track », *Text REtrieval Conference (TREC)*, National Institute of Standards and Technology (NIST), 2013.
- Eickhoff C., de Vries A. P., Collins-Thompson K., « Copulas for Information Retrieval », *In Proceedings of the 36th annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, Dublin, Ireland, 2013.
- Gerani S., Zhai C., Crestani F., « Score transformation in linear combination for multi-criteria relevance ranking », *In Proceedings of the 34th European Conference on Advances in Information Retrieval, ECIR'12*, Springer-Verlag, Berlin, Heidelberg, p. 256-267, 2012.
- Göker A., Myrhaug H., « Evaluation of a mobile information system in context », *Inf. Process. Manage.*, vol. 44, n° 1, p. 39-65, 2008.
- Grabisch M., « Fuzzy integral in multicriteria decision making », *Fuzzy Sets and Systems*, vol. 69, n° 3, p. 279-298, 1995.
- Grabisch M., Murofushi T., Sugeno M., Kacprzyk J., *Fuzzy Measures and Integrals. Theory and Applications*, Physica Verlag, Berlin, 2000.
- Mata F., Claramunt C., « GeoST : geographic, thematic and temporal information retrieval from heterogeneous web data sources », *In Proceedings of the 10th international conference on Web and wireless geographical information systems*, Springer-Verlag, Berlin, Heidelberg, p. 5-20, 2011.
- Moulahi B., Tamine L., Ben Yahia S., « L'intégrale de Choquet discrète pour l'agrégation de pertinence multidimensionnelle », *CORIA*, p. 399-414, 2013.
- Nagmoti R., Teredesai A., De Cock M., « Ranking Approaches for Microblog Search », *In Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, vol. 01 of *WI-IAT '10*, IEEE Computer Society, Washington, DC, USA, p. 153-157, 2010.
- Palacio D., Cabanac G., Sallaberry C., Hubert G., « On the evaluation of Geographic Information Retrieval systems : Evaluation framework and case study », *Int. J. Digit. Libr.*, vol. 11, n° 2, p. 91-109, 2010.
- Saracevic T., « Relevance : A review of the literature and a framework for thinking on the notion in information science », *Advances in Librarianship*, Academic Press, p. 79-138, 1976.
- Saracevic T., « Relevance : A review of the literature and a framework for thinking on the notion in information science. Part III : Behavior and effects of relevance », *Journal of the American Society for Information Science*, vol. 58, n° 13, p. 2126-2144, 2007.
- Sieg A., Mobasher B., Burke R., « Web search personalization with ontological user profiles », *In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, ACM, New York, NY, USA, p. 525-534, 2007.
- Tamine L., Bahsoun W., « Définition d'un profil multidimensionnel de l'utilisateur : Vers une technique basée sur l'interaction entre dimensions », *CORIA*, p. 225-236, 2006.
- Taylor A. R., Cool C., Belkin N. J., Amadio W. J., « Relationships between categories of relevance criteria and stage in task completion », *Information Processing and Management*, vol. 43, n° 4, p. 1071-1084, 2007.