Une Approche Possibiliste pour la Recherche d'Information

Asma H. BRINI (*) **Mohand BOUGHANEM** (*) **et Didier DUBOIS** (*)

{brini, bougha, dubois}@irit.fr

(*) IRIT, 118 route de Narbonne 31062 CEDEX 4, Toulouse, France.

Mots clefs :

Recherche d'Information, Réseaux Possibilistes, Réseaux Bayésiens.

Keywords:

Information Retrieval, Possibilistic Networks, Bayesian Networks

Palabras clave :

Escudriñar científico y tecnológico, administración del conocimiento, ingeniería del conocimiento, innovación, formalización del conocimiento, reunir de información

Résumé

Ce papier décrit une nouvelle approche pour un modèle de Recherche d'Information. Cette approche traduit à travers des réseaux Bayésiens naïfs des relations de dépendance entre les documents et les termes d'indexation quantifiables par deux mesures : la possibilité et la nécessité. Le processus de recherche restitue les documents plausiblement ou nécessairement pertinents à un besoin utilisateur. Ce besoin utilisateur est vu comme une nouvelle information à propager dans les réseaux.

1 Introduction

La problématique majeure de la Recherche d'Information (RI) consiste à extraire à partir d'une collection de documents, ceux qui répondent à un besoin utilisateur en se basant souvent sur des informations pauvres. Les différents modèles connus de la RI (booléen, vectoriel, probabiliste, bayésiens) représentent les documents et les requêtes sous forme de listes de termes pondérés puis mesurent une valeur de pertinence (similarité vectorielle, probabilité de pertinence) en se basant sur ces termes et leurs poids. La pondération des termes est à notre sens l'élément fondamental de tous les modèles de RI actuels [13], [16]. Lorsqu'elle est calculée automatiquement, cette pondération est obtenue à partir de combinaisons des fréquences d'apparition des termes dans les documents (tf), des fréquences d'apparition des termes dans la collection (idf) et de la longueur des documents (dl) [14] [18]. Quelque soit le modèle, la réponse à une requête est une liste de documents ordonnés selon cette valeur de pertinence. Certaines approches considèrent les poids des termes comme des probabilités de pertinence. Dans ces modèles, l'incomplétude (ou imprécision) de l'information n'est pas considérée lors de la représentation d'un document ou de son évaluation étant donnée une requête. En réalité, les notions de possibilité ou de certitude sont laissées en marge lors des calculs de la pertinence. Les méthodes actuelles utilisées pour représenter les documents (ensemble de termes et de leurs poids) ainsi que pour représenter le besoin utilisateur ne sont pas totalement comptatibles avec une définition précise de la pertinence. L'objectif de ce travail est donc de proposer une approche basée sur les mesures de nécessité et de possibilité pour un modèle de Recherche d'Information (RI). Un tel modèle devrait être capable de répondre à des propositions du type :

– il est plausible à un certain degré que le document constitue une bonne réponse à la requête,
– il est nécessaire, certain (dans le sens possibiliste), que le document répond à la requête,
– le document d_1 est préférable au document d_2 ou l'ensemble $fd_1; d_2g$ est préférable à l'ensemble $fd_3; d_4g$.

Le premier type de proposition vise à éliminer certains documents de la réponse (”weak plausibility”). La seconde réponse se focalise sur les documents qui seraient pertinents. Le dernier type de proposition suggère que puisque l'ordonnancement des documents en réponse à un besoin utilisateur peut être traité d'une manière qualitative, les approches ordinales pourraient être utilisées. La définition de la pertinence d'un document vis à vis d'une requête, en fonction des données dont nous disposons, est difficilement exprimable (ou traductible) par une unique mesure de probabilité. En effet, celle-ci ne tient pas compte des notions d'imprécision et de vague intrinsèque à la pertinence [4]. En réalité, une mesure de probabilité portant sur un événement et son contraire est quelque peu restrictive. Dans le modèle proposé, un document contenant tous les termes de la requête constitue une réponse possiblement pertinente à la requête. Cette plausibilité doit être renforcée par une certitude provenant de la mesure de nécessité. L'approche proposée est basée sur un réseau possibiliste, où les valeurs des liens reliant les nœuds documents aux nœuds termes sont obtenues par des mesures de nécessité et de possibilité. La mesure de possiblité est utile pour filtrer les documents et la mesure de nécessité pour renforcer la pertinence des documents restants.

L'utilisation des réseaux et plus particulièrement des réseaux Bayésiens est discutée en section 2. En section 3, des notions de la théorie des possibilités sont brièvement présentées. L'approche proposée est décrite en section 4 et appliquée dans un exemple en section 5. Finalement, la section 6 donne les perspectives de ce travail.

2 Réseaux Bayésiens en RI

Les Réseaux Bayésiens (RBs) [5], [10], [11] fournissent un formalisme flexible pour la combinaison d'information provenant de différentes sources. Lorsque les mesures de probabilité dépendent d'une vue subjective, elles ne traduisent pas forcément des fréquences relatives, mais fournissent un degré de croyance d'un événement et d'un événement par rapport à un autre événement (probabilité conditionnelle). Depuis 1990, les RBs ont été utilisés en RI. Ils fournissent un formalisme pour combiner des informations provenant de différentes sources (requêtes du passé, réinjection de pertinence), pour restituer les documents, et ont permis de combiner différentes approches de RI [12]. Les modèles les plus connus en RI utilisant les RBs sont les Réseaux d'Inférence et les Réseaux de Croyance. Les réseaux d'Inférence sont utilisés dans le système INQUERY [19] [20] et ses performances sont liées à sa capacité à représenter différentes approches de la RI et à les combiner dans un seul modèle. Le réseau d'inférence est composé de deux réseaux : le réseau document et le réseau requête. Le réseau document représente les documents de la collection et contient différents schémas de représentation (résumés, textes, etc). Les nœuds du réseau requête représentent les concepts de la requête et le besoin utilisateur. Les réseaux document et requête sont liés par l'intermédiaire des nœuds termes d'indexation. Les valeurs des nœuds sont binaires {vrai, faux} et les valeurs des arcs reliant les nœuds termes au nœud requête sont obtenues par l'utilisation d'un des schémas des modèles connus de la RI (booléen, vectoriel etc). Ce système évalue la pertinence du document étant donnée une requête, le résultat est une liste de documents pondérés. Ces poids sont considérés comme un coefficient de similarité proportionnel à la fréquence des termes dans le document et inversement proportionnel à celle dans la collection. D'autres travaux basés sur ces réseaux ont été proposés pour les systèmes hypertextes [15]. Les Réseaux de Croyance (RC) [12], [17] ont été utilisés pour dériver des connaissances des requêtes du passé et les combiner avec le modèle vectoriel [12]. La sélection d'un document est basée sur la similarité entre le document d_j et la requête Q, calculant la probabilité $P(d_j = 1/Q = 1)$. $Q = 1$ et $d_j = 1$ signifient respectivement Q activé et d_j activé. Crestani & al [6], ont proposé un modèle pour la RI basé sur les réseaux Bayésiens pour les documents structurés. Un réseau à deux structures (BNR- 2) [7] a été conçu et étendu à un réseau multi-structures. L'ensemble des variables dans le modèle BNR-2 est composé de deux ensembles distincts, l'ensemble des variables aléatoires binaires définissant les termes du dictionnaire et l'ensemble des variables aléatoires binaires représentant les documents de la collection. Chaque document est composé d'une structure hiérarchique de différents niveaux d'abstraction (titre, auteur, section, paragraphe etc). Le processus d'inférence calcule, étant donnée une requête, les probabilités a priori de la pertinence de toutes les unités de structure. Les documents de score élevé sont restitués. Certaines récentes recherches [7] [8] ont proposé des modèles de Réseaux Bayésiens avec une topologie flexible qui peut tenir compte des relations de dépendance existant entre les termes ou les documents. Le sens des représentations des documents et du besoin utilisateur pour tous ces modèles est identique. Cependant, le modèle proposé ici, tente de fournir un autre sens à ces représentations ainsi qu'à l'évaluation (comparaison de ces deux représentations). Une manière de répondre à la problématique peut être apportée par l'utilisation des Réseaux Possibilistes (RP).

3 Logique Possibiliste

La théorie des possibilités introduite par Zadeh [22] et développée par Dubois et Prade [9] traite l'incertitude sur l'intervalle [0,1], appelé échelle possibiliste, d'une manière qualitative ou quantitative. Nous nous restreignons, pour cette première approche, au cadre quantitatif.

3.1 La théorie des Possibilités

Distribution de possibilité

La théorie des possibilités est basée sur les distributions de possibilité. Cette dernière, notée par π est une application de Ω (l'univers de discours) vers l'échelle traduisant une connaissance partielle sur le monde. L'échelle possibiliste est définie de deux manières. Dans le cadre numérique les valeurs des possibilités traduisent souvent les bornes supérieures des probabilités. La combinaison des distributions de possibilité, exprimée à l'aide des normes triangulaires (t-normes) dépendent du cadre. Dans le cadre numérique, l'opérateur « produit » peut être utilisé pour combiner des distributions de possibilité indépendantes.

Mesures de Nécessité et de Possibilité

Dire qu'un événement est non possible n'implique pas seulement que son événement contraire est possible mais qu'il est certain. Deux mesures duales sont utilisées : la mesure de possibilité $\Pi(\phi)$, et la mesure de nécessité $N(\phi)$.

3.2 Conditionnement Possibiliste

En logique possibiliste, le conditionnement consiste à modifier la distribution de possibilité initiale π à l'arrivée d'une nouvelle information *i*. Soit ϕ, une sous classe de Ω, ϕ = [i] l'ensemble des modèles de i. La distribution initiale π est remplacée par π'=π(•/ϕ). Dans un cadre quantitatif, les éléments de ϕ sont proportionnellement modifiés :

$$\pi(\omega /_p \phi) = \begin{cases} \dfrac{\pi(w)}{\Pi(\phi)} & si \ \ w \in \phi \\ 0 & \text{sin } on \end{cases} \quad [1]$$

avec
$/_p$: pour désigner le conditionnement dans un cadre quantitatif.

3.3 Réseaux Possibilistes (RP)

Les travaux existants sur les réseaux possibilistes sont soit des adaptations directes de l'approche probabiliste [2], ou des méthodes d'apprentissage à partir de données imprécises [3]. La théorie des possibilités offre deux définitions du conditionnement, ce qui conduit à deux définitions des réseaux causaux possibilistes. Les réseaux possibilistes basés sur le produit sont très similaires aux réseaux probabilistes.

3.3.1 Définitions

Un graphe possibiliste orienté sur un ensemble de variables *V* est caractérisé par une composante qualitative et une composante numérique. La première est un graphe acyclique orienté. La structure du graphe représente l'ensemble des relations d'indépendance. La seconde composante quantifie les liens du graphe en utilisant des distributions de possibilité conditionnelles de chaque nœud dans le contexte de ses parents. Ces distributions de possibilité doivent vérifier la contrainte de normalisation. Pour chaque variable V :
– Si *V* est un nœud racine et D_V le domaine de *V*, la possibilité a priori de *V* doit satisfaire :

$$\max_v \Pi(v) = 1, \forall v \in D_v$$

– Si *V* n'est pas un nœud racine, la distribution conditionnelle de *V* dans le contexte de ses parents doit satisfaire :

$$\max_v \prod(v / Par_v) = 1, \forall v \in D_v, Par_v \in D_{Par_V}$$

avec :
D_V : le domaine de V ,
Par_V : l'ensemble des parents de V ,
D_{Par_V} : le domaine des parents de V .

3.3.2 Réseaux possiblistes basés sur le produit

Un graphe possibiliste basé sur le produit, noté par GP_P , est un graphe possibiliste où les possibilités conditionnelles sont obtenues par le conditionnement produit. La distribution de possibilité des réseaux possibilistes basés sur le produit, notée par π_p, est obtenue par la règle de chaînage :

$$\pi_p(V_1,.....,V_N) = PROD_{i=1,...,N}(V_i / Par_{V_i})$$

avec :
PROD : l'opérateur produit.

4 Un modèle Possibiliste pour la RI

Le modèle proposé utilise d'une nouvelle manière les connaissances disponibles. Ces connaissances concernent les documents de la collection ainsi que la liste des termes d'indexation et de leur fréquence. Les documents de la collection ainsi que leurs termes d'indexation sont représentés par des réseaux naïfs possibilistes. Considérant un terme relatif à un document, une relation de dépendance quantifiable existe entre un terme et un document. La requête déclenche un processus de propagation entraînant le changement de croyance sur les nœuds documents. Ce processus de recherche peut être analogue à une étape de diagnostic dans le domaine médical. La collection de documents est comme un ensemble de maladies possibles, les symptômes sont les termes. La requête est vue comme une observation. Le but étant de trouver la maladie plausiblement développée en observant le patient (requête), étant donnés les symptômes qu'il présente. Dans cet article la pertinence est représentée dans un seul cadre : le cadre quantitatif.

4.1 Architecture du modèle

Le modèle est représenté par un réseau possibiliste d'architecture définie dans la Figure (1) ci-dessous :

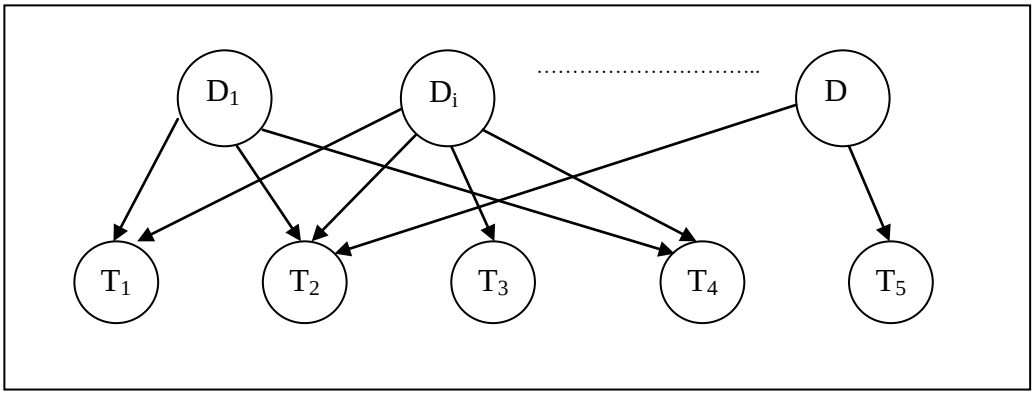

Figure 1 : Architecture générale du modèle

Avec :
Nœud D_j : nœud d'un document de la collection. Le domaine de D_j, noté dom(D_j), est binaire $\{d_j, \neg dj\}$. Une instanciation possible, $D_j = d_j$ signifie que le document D_j est pertinent. $D_j = \neg dj$, signifie que le document D_j est non pertinent.
Nœud T_i : nœud terme est un terme d'indexation du document. Les variables T_i sont binaires. Le domaine d'un terme, noté dom(T_i), est dom(T_i) $=\{t_i, \neg t_i\}$. $Ti = t_i$ signifie que le terme i est représentatif du document, $Ti = \neg t_i$ signifie que le terme i est non représentatif du document. Ce domaine est lié au contexte du parent.
Arc : un arc orienté d'un nœud document vers les nœuds termes d'indexation exprime une relation de dépendance entre le document et les termes qu'il contient. Un arc entre un nœud T_i et un nœud D_j traduit la possibilité et la nécessité que T_i soit représentatif (ou non) du document D_j et ceci en fonction de sa fréquence dans le document et de celle dans la collection.

4.2 Poids

Pour évaluer la possibilité et la nécessité de pertinence, nous avons besoin de définir explicitement la pertinence représentée par des arcs dans le réseau. Une nouvelle interprétation de la pondération des termes est suggérée. L'approche proposée tente de distinguer entre les termes possiblement représentatifs des documents (ceux qui sont absents sont écartés) et ceux nécessairement représentatifs, c'est-à-dire les termes qui suffisent à caractériser les documents.
H*ypothèse 1 :* Un terme est d'autant moins représentatif d'un document qu'il apparaît peu fréquemment dans ce document ;
Hypothèse 2 : Un terme est d'autant plus nécessairement représentatif du document qu'il apparaît fréquemment dans ce document et peu fréquemment dans les autres documents de la collection.
Donc d'après l'hypothèse 1, $\prod(t_i/d_j)$ peut être estimée par :

$$\prod(t_i / d_j) = nft_{ij}$$

Où :

nft_{ij} la fréquence normalisée : $nft_{ij} = \dfrac{tf_{ij}}{\max_{\forall t_k \in d_j} tf_{kj}}$

Un terme de poids 0 signifie que le terme n'est pas compatible avec le document. S'il est égal à 1, alors le terme est possiblement représentatif ou pertinent pour décrire (donc représenter) le document. Ici, le terme « représentatif » ne doit pas être considéré au sens large, mais comme « pertinent pour restituer le document ». Si un terme est représentatif du document, dans le sens général, il n'aiderait pas forcément à restituer le document. Typiquement, pour un document traitant de la « logique floue », le terme « floue » est très représentatif, mais uniquement potentiellement, puisqu'il ne le caractérise pas sur une collection de documents traitant du même domaine. Notons que le degré de possibilité est normalisé (son maximum vaut 1). Ce degré évalue à quel point un terme est "typique" du document et donc à quel point il est possible qu'il contribue à sa restitution. Un terme qui n'apparaît pas dans un document est considéré comme non compatible avec le document et s'il apparaît avec une fréquence maximale, alors il est considéré comme un candidat possible à sa représentation. En logique possibiliste, la mesure de possibilité possède une mesure duale : la nécessité. Celle-ci, dans ce

contexte, exprime l'idée que s'il est certain qu'un terme ne représente pas un document, alors il est certain que le terme n'implique pas le document. Cette certitude est exprimée par :

$$N\left(d \rightarrow \overline{t_i}\right) \geq 1 - nft_{ij} [3]$$

Un terme discriminant dans une collection, est un terme qui apparaît fréquemment dans peu de documents de la collection. Un terme discriminant est un terme nécessairement représentatif du document, il contribue à sa sélection et donc à sa restitution en réponse à une requête. Nous définissons, un degré de nécessaire pertinence, ϕ_{ij}, du terme i pour représenter le document j par :

$$N\left(d \rightarrow \overline{t_i}\right) \geq \phi_{ij} \qquad [4]$$

$$\phi_{ij} = \mu_1\left(\frac{nC}{nd_i}\right) x \mu_2\left(nft_{ij}\right)$$

Où :
nC : nombre de documents de la collection,
nd_i : nombre de documents contenant le terme i,
μ_1, μ_2 : fonctions de normalisation.
Par exemple :
μ_1: une t-norme telle que la fonction logarithmique,
μ_2 : la fonction identité.
Ce degré de nécessaire pertinence va donc permettre de calculer la possiblité que le terme n'implique pas le document par :

$$\Pi\left(t_i \wedge \overline{d_j}\right) \leq 1 - \phi_{ij} \qquad [5]$$

Soit la distribution de possibilité définie (cf. tableau 1) sur :

$$\{d_j, \neg d_j\} x \{t_i, \neg t_i\},$$

$d_j, \neg d_j$: un document instancié signifie respectivement que le document est pertinent, non pertinent, étant donnée une requête,
$t_i, \neg t_i$: un terme instancié signifie respectivement que le terme est représentatif, non représentatif d'un document.

	d_j	$\neg d_j$
t_i	nft_{ij}	$1-\phi_{ij}$
$\neg t_i$	1	1

Tableau 1 – Distribution de possibilité.

Le tableau 1 ci-dessus, donne la distribution de possibilité la moins spécifique obéissant aux contraintes (4) et (5). Notons par ailleurs que :

$$\Pi(d_j) = \Pi(\overline{d_j}) = 1$$
$$\Rightarrow \Pi(t_i / d_j) = \Pi(t_i \wedge d_j) = nft_{ij}$$

4.3 Un simple schéma de propagation

Dans le cadre numérique, les valeurs de possibilité et de nécessité, a priori et conditionnelles, ont un sens. L'idée est de répondre à des propositions du type :

– « d_1 est pertinent pour Q » est possible ou non, quantifiée par $\Pi(d_1/Q)$
– « d_1 est pertinent à Q » est certain ou non, quantifiée par $N(d_1/Q)$

Pour le modèle de base présenté ici, la requête est composée de simples mots clés. Lorsque la requête est connue, un processus de propagation est déclenché à travers le réseau, modifiant les valeurs des possibilités a priori reliant les documents aux termes d'indexation. Dans ce modèle, la formule de propagation est identique à celle des réseaux Bayésiens naïfs [2]. Cependant, deux évaluations sont réalisées : $\Pi(d_1/Q)$ et $\Pi(\neg d_1/Q)$ (car leur somme ne vaut pas 1).

Soit une requête $Q = (t_1, \ldots, t_T)$ interprétée conjonctivement,

$$\Pi(d_i / Q) = \frac{\Pi(Q / d_i)\Pi(d_i)}{\Pi(Q)} \qquad [7]$$

La possibilité de pertinence évalue à quel point $D_i = d_i$ est possiblement pertinent étant donnée une requête *Q*. Lorsque cette valeur vaut 0 le document est écarté. On suppose l'indépendance des termes.

H*ypothèse 3 :* pour chaque document de la collection, les termes sont conditionnellement indépendants dans un document donné. Si le document D_i est composé des termes *T*, l'hypothèse ci-dessus transforme la formule (7) lorsque le document est instancié ($D_i = d_i$) :

$$\Pi'(d_i / Q) = \frac{\Pi(t_1 / d_i) * .... * \Pi(t_T / d_i) * \Pi(d_i)}{\Pi(Q)} \qquad [8]$$

Pour comparer les possibilités de pertinence des documents de la collection, uniquement le numérateur est utile. Le numérateur de la formule (8) mesure la pertinence potentielle d'un document étant donnée une requête.

La certitude de restituer un document étant donnée une requête, notée *N(dj=Q)*, est donnée par :

$$N(d_{ji} / Q) = 1 - \Pi(d_j / Q),$$
$$\Pi(\neg d_j / Q) = \frac{\Pi(Q / \neg d_i)\Pi(\neg d_i)}{\Pi(Q)} \qquad [10]$$

Lorsque le document est instancié et d'après l'hypothèse 3, nous pouvons écrire :

$$\Pi'(\neg d_j / Q) = \frac{\Pi(t_1 / \neg d_i) * .... * \Pi(t_T / \neg d_i) * \Pi(\neg d_i)}{\Pi(Q)} \qquad [11]$$

Le numérateur peut être exprimé par :

$$\Pi''(\neg d_j / Q) = (1 - \phi_{1j}) * ..... * (1 - \phi_{Tj}) \qquad [12]$$

Les documents préférés sont ceux qui ont une valeur $N(d_j/Q)$ élevée parmi ceux qui ont une valeur $\Pi''(d_i=Q)$ élevée aussi. Si $N(d_j/Q)$ vaut zéro, les documents restitués sont ceux qui ont une valeur $\Pi''(d_i=Q)$ élevée.

5 Exemple

Soit une collection de 3 documents :
$d_1 = \{t_1, t_1, t_1, t_2, t_2, t_3\}$,
$d_2 = \{t_1, t_1, t_2, t_2, t_2, t_2\}$,
$d_3 = \{t_1, t_3, t_3, t_3, t_3, t_4, t_4\}$.
Les matrices des liens de ces trois documents sont données dans le tableau ci-dessous (pour les documents D_2 et D_3 nous ne donnons que la ligne de la représentativité).

	d_i			$\neg d_i$		
	t_1	t_2	t_4	t_1	t_2	t_4
Représentativité (i=1)	1	2/3	0	1	0.88	1
Non représentativité (i=1)	1	1	1	1	1	1
Représentativité (i=2)	½	1	0	1	0.82	1
Représentativité (i=3)	¼	0	½	1	1	0.76

Tableau 2 – Matrice des liens des documents D_1, D_2, D_3

Soit une requête Q= $\{t_1, t_2, t_4\}$.
Cette requête interprétée comme une conjonction de termes serait trop restrictive puisque aucun document de la collection ne contient les trois termes à la fois. La nécessité et la possibilité d'avoir un des documents de cette collection en résultat sont nulles. Pour éviter d'obtenir une liste de documents vide en résultat, nous cherchons les documents qui contiennent au moins deux termes de la requête puis au moins un terme (si aucun document de la collection ne contient deux termes) ; dans ce cas, la possibilité de tous les documents vaut 1 et leur nécessité vaudra 0. Nous cherchons alors les documents qui traitent des termes $\{t_1, t_2\}$, $\{t_1, t_4\}$, $\{t_2, t_4\}$. Nous voyons à travers cet exemple, la nécessité de permettre à l'utilisateur d'exprimer des préférences entre les termes de la requête. Imaginons la requête Q'=$\{(t_1, t_4) > t_2\}$ exprimant une préférence des termes $\{t_1, t_4\}$. Sachant qu'aucun document ne contient les trois termes à la fois, le processus d'évaluation calcule la requête
Q'=$\{t_1, t_4, \neg t_2\}$

$$\Pi(dj/Q) = \max\left(\Pi(t_1/dj) * \Pi(\neg t_2/d_j) * \Pi(t_4/d_1)\right)$$

L'évaluation des documents d_1; d_2; d_3 pour cette requête donne :
$\Pi(d_1/Q'') = \Pi(d_2/Q'') = 0, \Pi(d_3/Q'') = 1/8$
Soit
$\Pi(\neg d_1/Q'') = \Pi\neg(d_2/Q'') = 1,$
$N(d_1/Q'') = N(d_2/Q'') = 0,$
$\Pi(\neg d_3/Q'') = 0.76, N(d_3/Q) = 0.24$

On ignore totalement si d_1 et d_2 sont pertinents. Nous remarquons pour le document d_3 que la nécessité est supérieure à 0 mais que la possibilité ne vaut pas 1. Cela indique que les termes choisis tendent à sélectionner ce document, même si ces termes ne sont pas les plus fréquents du document. Pour la requête Q'' le document d_3 est préféré aux documents d_1 et d_2.

6 Conclusion

Ce papier propose les étapes préliminaires pour un modèle de RI en utilisant la logique possibiliste. L'approche proposée fournit un nouveau cadre pour l'évaluation de la pertinence aussi bien pour la représentation des documents et de la requête que pour la sélection des documents en réponse à un besoin utilisateur, et ceci en modélisant l'imprécision et le vague dans la définition de la pertinence. Les mesures de possibilité et de nécessité sont utilisées pour quantifier les relations de dépendance (ou indépendance) entre les termes et les documents qu'ils indexent et pour restituer les documents nécessairement ou possiblement pertinents étant donné une requête. Les expérimentations sont en cours d'évaluation sur la collection TREC 10 [21]. Comme perspectives, nous prévoyons de considérer les préférences entre les termes de la requête mais aussi entre les documents restitués. De plus, l'approche de base tient uniquement compte de l'aspect quantitatif, nous tentons de l'étendre au cadre qualitatif. possibiliste pour montrer comment, dans notre cas particulier, elle peut généraliser les réseaux Bayésiens probabilistes.

7 Bibliographie

[1] B. Yates, R and Ribeiro Neto, B *Modern Information Retrieval*, Addison Wesley, 1999.

[2] S. Benferhat, D. Dubois, L.Garcia, H. Prade (19). Possibilistic Logic Bases and Possibilistic Graphs. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI'99)*, pages 57-64, 1999.

[3] C. Borgelt, J. Gebhardt, and R. Kruse. Possibilistic Graphical Models. *Computational Intelligence in Data Mining*, pp. 51-68. CISM Courses and Lectures 408. 2000.

[4] A. H. Brini, M. Boughanem. Relevance feedback introduction of partial assessments for query expansion. EUSFLAT, p. 67-72, 2003.

[5] W. Buntine (1994). Representing Learning with graphical Models.*Technical Report*, FIA 94-14, Artificial Intelligence Research Branch, NASA Ames Research Center, 1994.

[6] Crestani, F., Luis M. de Campos, Juan M. Fernandez-Luna, Juan F. Huete. A Multi-layered Bayesian Network Model for Structured Document Retrieval. *ECSQARU 2003 :* 74-86. 2003.

[7] De Campos, L.M., Fernandez-Luna, Juan M, Huete, Juan F. (2003). The BNR Model : foundations and performance of Bayesian Network-based retrieval model. *JASIST*,54(4),302-313, 2003.

[8] De Campos, L.M., Fernandez-Luna, Juan M, Huete, Juan F.(2003). Two Term-Layers : an Alternative Topology for Representing Term Relationships in the Bayesian Network Retrieval Mode. Advances in Soft Computing-Engineering, Design and Manufacturing(pp. 213-224). 2003.

[9] D. Dubois, H. Prade (1998). Possibility Theory :Qualitative and Quantitative Aspect. Handbook on Defeasible Reasonning and Uncertainty Managment Systems, volume 1, pages 21-42. 1998.

[10] Finn V. Jensen (2000). *Bayesian Networks and Decision Graphs*, 2000. [11] Judea Pearl (1988), *Probabilistic Reasonning in Intelligent Systems : Networks of Plausible Inference*. 1988.

[12] B. Ribeiro-Neto, I. Silva, R. Muntz. A Belief Network Model for IR. In *Proc. Of the 19th ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 253-260, Zurich. 1996.

[13] Robertson, S.E., Walker, S., Hancock-Beaulieu,(1994) M.M. : Okapi at TREC 3. In *Proceedings of the 3rd Text Retrieval conference (TREC 3)*, pages 109-126, NIST. 1994.

[14] G. Salton, J. Allan, C. Buckley and A. Singhal : Automatic Analysis, Theme Generation and Summarization of Machine Readable Texts. *Science*, 264, 3, 1421-1426. 1994.

[15] Jacques Savoy, Daniel Desbois : Information Retrieval in Hypertext Systems an Approach Using Bayesian Networks. Electronic Publishing, Vol. 4(2), 87-108, 1991.

[16] K. Sparck Jones : A Look Back and a Look Forward, *Proceedings of the 11th annual international ACM SIGIR conference on Research and Development in Information Retrieval*, p.13-29. 1988.

[17] Ilmerio Silva, Berthier Ribeiro-Neto, Pavel Calado, Edleno Moura, and Nivio Ziviani : Link-Based and Content-Based Evidential Information in a Belief Network Model. *In ACM SIGIR* 23rd Int. Conference on Information Retrieval, pages 96-103. 2000.

[18] A. Singhal, G Salton, M. Mitra, C. Buckley : Document Length Normalization. *Information Processing and Managment (IPM)*, 32(5) : 619-633. 1996.

[19] H.R. Turtle and W.B. Croft (1990). Inference networks for document retrieval. In *Proceedings of the 13th International Conference on Research and Development in Information Retrieval*, pp 1-24, 1990.

[20] H.R. Turtle (1991). *Inference Networks for Document Retrieval*, Thesis, University of Massachussets. 1991.

[21] E. M. Voorhees. Overview of the TREC 2001 question answering track. *Proceedings of the Tenth Text REtrieval Conference (TREC 2001)*, 2001.

[22] L. A. Zadeh. Fuzzy Sets as a Basis for a theory of Possibility. In *Fuzzy Sets and Systems*, 1 :3-28, 1978.