

---

## Recherche de documents structurés en mobilité : un modèle et une mesure d'évaluation.

**Philippe Mulhem, Delphine Verbyst**

Laboratoire LIG-CNRS  
Équipe MRIM - Bât B  
Domaine Universitaire  
385 rue de la Bibliothèque  
F-38400 Saint Martin d'Hères  
Philippe.Mulhem@imag.fr, Delphine.Verbyst@imag.fr

---

*RÉSUMÉ. Les travaux en recherche de documents structurés ne s'intéressent que peu pour le moment à leur emploi dans un contexte de mobilité. Dans de tels cas, on pose que l'objectif d'un système de recherche d'information demeure inchangé (il est de satisfaire le besoin d'information d'un utilisateur), mais il faut intégrer que les écrans des systèmes mobiles sont petits et que le réseau utilisé n'a pas une grande bande passante. Nous proposons dans cet article une première approche, relativement simple, qui prend en compte ces aspects, en privilégiant les parties de documents structurés petites en terme de taille. Pour tout travail scientifique, il est nécessaire d'évaluer les résultats de nos propositions, c'est pour cela que nous proposons une mesure d'évaluation des systèmes de recherche d'information en mobilité. Cette mesure est inspirée du gain cumulé déjà utilisé pour les documents structurés, mais intègre la taille des documents pour en calculer leur intérêt dans un contexte de mobilité en favorisant les parties de documents petites.*

*ABSTRACT. Research works on structured documents only focus a bit on the context of mobility. In such a case, we assume that the objectives of an information retrieval remain the same: we need to satisfy the user information needs. We need however to take into account the sizes of the display devices and the bandwidth of the networks used. We propose here a first attempt that considers such constraints by putting emphasize on small document parts. As for any scientific work, evaluation is a must. That is why we propose a way to include such constraints when evaluating information retrieval systems in a mobile context. The proposed measure comes from cumulated gain values, already in use for structured document retrieval, and includes the document size when computing the relevance of documents in a mobile context.*

*MOTS-CLÉS: contexte matériel, gain cumulé*

*KEYWORDS: hardware context, cumulated gain.*

---

## Introduction

L'objectif de cet article est double : tout d'abord caractériser les besoins d'un système de recherche d'information en mobilité, et ensuite proposer un moyen d'évaluer la qualité de tels systèmes.

Avant d'aller plus loin, nous émettons tout d'abord l'hypothèse qu'un système prenant en compte le contexte de mobilité est sensé fournir de meilleurs résultats : "The underlying hypothesis (and belief) is that by taking account of context the next generation of retrieval engines dependent on models of context can be created, designed and developed delivering performance exceeding that of out-of-context engines." [Ingwersen & Järvelin 2005].

Un système de recherche d'information en mobilité nécessite, de par ses contraintes (taille d'écran, débit réseau), de favoriser des résultats courts. Un document long en réponse a comme premier désavantage de ne pas pouvoir être visualisé dans sa totalité en une fois sur un petit écran, et comme second désavantage de prendre du temps pour être rapatrié sur le dispositif mobile. Un moyen d'aborder ce problème est de se situer dans un cadre de recherche de documents structurés et de considérer que le système évaluera la pertinence des parties de ces documents (appelées *doxel* pour *document element*), ces parties étant plus petites que les documents complets. Pour les parties de documents pertinentes, plus elles sont petites et plus il sera facile de les rapatrier et de les afficher sur le matériel mobile.

Proposer un modèle pour de tels systèmes est donc utile, mais il est difficile de se faire une idée de la qualité des résultats obtenus si l'on ne dispose pas d'un outil d'évaluation adapté. Nous proposons donc également ici un moyen d'évaluer la capacité d'un système de recherche d'information à retourner en réponse des documents petits. Nous nous basons pour cela sur une extension des calculs de gain cumulé, déjà utilisés dans la compétition INEX 2006 [Lalmas et al. 2006]. Dans ce cadre, nous ne proposons pas d'évaluer les systèmes en mobilité avec des utilisateurs réels, ce qui est très difficile à contrôler et qui dépasse la cadre de cet article, mais de rester dans l'approche courante de la recherche d'information en définissant une collection de test avec un corpus de documents structurés, et un ensemble de requêtes résolues.

Cet article est organisé de la manière suivante. En partie 1, nous proposons un bref état de l'art sur les approches actuelles pour la recherche d'information en mobilité. Nous proposons ensuite une description détaillée des mesures de gain cumulé telles qu'elles sont utilisées dans la compétition INEX. Nous proposons en partie 3 un modèle de recherche de documents en contexte de mobilité, avant de proposer une mesure pour évaluer de tels systèmes. Une évaluation de nos propositions est réalisée en partie 5. Nous concluons en partie 6 en indiquant les directions futures de ce travail.

## 1. Recherche d'information en contexte de mobilité

Pour la recherche de documents en contexte de mobilité, nous nous intéressons ici aux aspects interaction entre l'utilisateur et le système.

Si nous considérons l'interface par laquelle l'utilisateur peut exprimer son besoin d'information, le fait d'être en contexte de mobilité a un impact. Par exemple, le travail de Du et Crestani dans [Du & Crestani 2006] a montré que l'expression des requêtes écrites ou parlées diffère, en particulier au niveau du nombre de termes : 23 en moyenne pour les requêtes parlées et 7 termes pour les requêtes écrites. Ces résultats obtenus sur la langue anglaise sont cohérents avec d'autres résultats des mêmes auteurs pour le chinois mandarin [Du & Crestani 2004]. Un système ne donnera pas les mêmes résultats à des besoins d'informations similaires suivant le mode d'interaction en entrée.

Un système de recherche d'information en mobilité doit également intégrer les capacités de sortie du système physique utilisé : le travail décrit dans [Milic-Frayling et al. 2003] utilise un affichage des dix premiers résultats de requête adapté, ainsi qu'une présentation de l'imagette d'une page avec laquelle l'utilisateur peut interagir, [Yin & Lee 04] ont proposé d'utiliser les liens entre les pages pour déterminer les éléments importants à visualiser, en sachant que sur un petit écran on ne peut pas tout afficher.

Nous considérons que le système de recherche d'information doit être en mesure d'intégrer les points liés au contexte de mobilité à un niveau de modélisation. Dans cet article nous nous focalisons sur les résultats des requêtes qui doivent être adaptés au matériel utilisé en s'intéressant particulièrement à la taille des doxels renvoyés. Cela est en phase avec les travaux qui s'intéressent à la capacité d'affichage des systèmes utilisés. Dès lors que ce point sur la taille des documents est intégré, il faut être en mesure d'en tenir compte quand on va évaluer les résultats des systèmes sensés répondre à un besoin de recherche en contexte de mobilité avec des petits écrans en particulier.

## 2. Mesures de gain cumulé

Les mesures courantes de rappel et de précision ne sont pas adaptées à des documents structurés car elles considèrent chaque document comme atomique et utilisent des valeurs de pertinences binaires. Dans notre cas, nous devons étudier des mesures adaptées aux documents structurés et intégrant le fait que la pertinence d'un document doit également tenir compte de sa taille. Nous basons notre proposition sur un existant qui est la campagne INEX<sup>1</sup> qui a débuté en 2002 et qui a depuis été reconduite jusqu'en 2007. INEX a pour objectif d'évaluer les systèmes de recherche d'informations de documents XML. Pour ces campagnes, un certain nombre de

---

<sup>1</sup> INEX : <http://inex.is.informatik.uni-duisburg.de/>

mesures ont été proposées au cours des années. Nous décrivons donc ici certaines de ces mesures, tout en insistant sur le fait que ces évaluations sont adaptées aux documents structurés mais pas dédiées à être appliquées au contexte de mobilité. Dans les campagnes 2006 et 2007 d'INEX, le corpus est composé d'environ 600000 documents XML provenant de Wikipedia<sup>2</sup> en anglais. Les requêtes sur 2006 et 2007 atteignent plus de 200 requêtes résolues, les ensembles de documents et parties de documents pertinents sont évalués sur l'union des résultats de tous les systèmes participants à la campagne. Parmi les nombreuses tâches de recherche que se propose d'évaluer INEX, nous nous intéressons à la tâche *Focused* lors de la campagne 2006. Dans cette tâche, l'idée est d'évaluer des réponses sous la forme d'une liste de doxels, sans organiser davantage les réponses ; un système retourne donc une liste de doxels qu'il juge pertinents, indépendamment de leur contexte d'occurrence. Pour ces évaluations, la liste des pertinents est composée de doxels qui ne se superposent pas, et pour chaque doxel pertinent il est stocké le nombre de caractères de la zone marquée explicitement pertinente par les évaluateurs humains. A partir de ces données, une mesure qui a été définie est le gain cumulé : les approches à base de gain cumulé intègrent des valeurs de pertinence non binaires, ce qui est important dans le cadre de ce travail. Les approches à base de gain cumulé se basent sur une indication du gain pour l'utilisateur lors de la visualisation d'un document retourné par un système. Pour calculer le gain cumulé d'une liste de résultats qui ne se superposent pas, un document en position  $i$  est remplacé par sa valeur de gain  $xG[i]$ . A un rang  $i$ , la mesure de gain cumulé de la liste, noté  $xCG[i]$  est la somme des gains de ses documents, comme le montre la formule [1].

$$xCG[i] = \sum_{j=1}^i xG[j] \quad [1]$$

La valeur  $xG[j]$  est calculée par la fonction  $spec(d_i)$ , avec  $d_i$  le doxel en position  $i$  de la liste considérée. Cette fonction  $spec$  renvoie la proportion de  $d_i$  qui est pertinente, en se basant sur la taille de  $d_i$  en nombre de caractères et sur la taille de la partie pertinente de ce doxel.

Pour normaliser les résultats des différents systèmes évalués, on définit une liste idéale  $LI$ , triée par ordre de pertinence décroissante. La valeur de pertinence d'un élément  $LI[k]$ , notée  $xI[k]$  est en fait le résultat de la fonction  $spec$  vue plus haut pour le doxel en position  $k$  de  $LI$ . Un gain cumulé  $xCI[k]$  sur cette liste est également calculé par la formule [2].

$$xCI[i] = \sum_{j=1}^i xI[j] \quad [2]$$

---

<sup>2</sup> Wikipedia, the free encyclopedia, [http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page)

La valeur de gain cumulé normalisée, noté  $nxCG[i]$  pour une position  $i$  est définie par la formule [3].

$$nxCG[i] = \frac{xCG[i]}{xCI[i]} \quad [3]$$

Cette mesure est intéressante car elle intègre le fait que chaque doxel n'a pas la même valeur de pertinence. Les recouvrements entre doxels pertinents et doxels renvoyés par un système sont traités par une base de rappel totale (*full recall-base*) qui contient pour chaque doxel contenant au moins une partie pertinente, le nombre de caractères pertinents permettant de calculer la fonction *spec* décrite plus haut.

Dans cet article, nous proposons d'intégrer la taille des documents dans le calcul de pertinence des documents, car dans un contexte de mobilité nous considérons que les documents petits sont les plus faciles à lire pour un utilisateur.

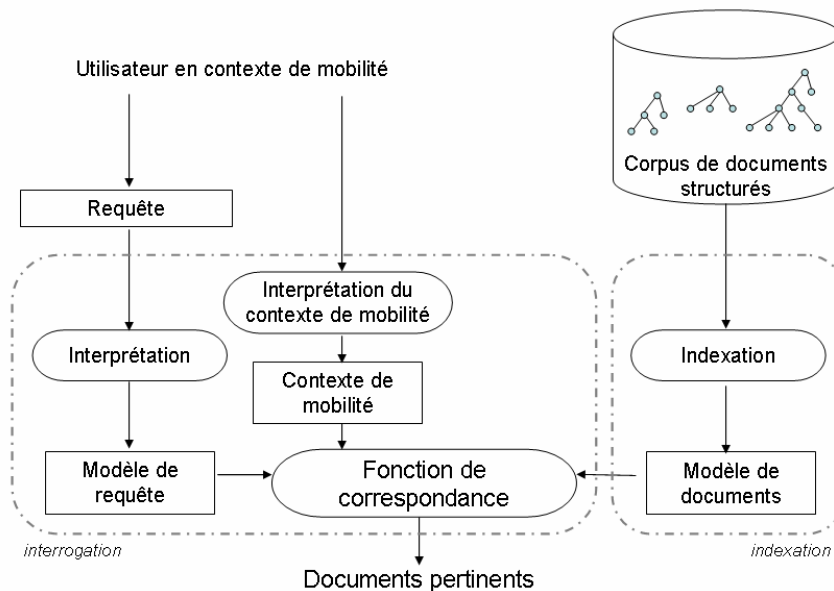


Figure 1 - Système de recherche de documents structurés en mobilité.

### 3. Modèle de recherche de documents structurés en mobilité

La figure 1 représente un système de recherche de documents structurés en mobilité. Cette décomposition fonctionnelle comporte essentiellement deux parties :

- Une partie « indexation », qui correspond à un traitement des documents structurés pour les insérer dans la base de données d'interrogation, sur laquelle porteront les requêtes des utilisateurs ;
- Une partie « interrogation », qui correspond à la partie gérant l'interaction avec l'utilisateur en contexte de mobilité, incluant le traitement de la requête et l'interprétation du contexte matériel, et qui renvoie les doxels pertinents.

### 3.1. Indexation des doxels

La représentation du contenu d'un doxel  $d_i$  est un vecteur généré à partir du modèle vectoriel usuel et qui porte sur la totalité du contenu du doxel:  $d_i = (w_{i,1}, \dots, w_{i,k})$ . Une telle représentation produit de bons résultats pour les documents structurés [Huang et al. 2006]. La pondération que nous utilisons est un simple *tf.idf*, avec un *idf* basé sur le corpus entier et avec les normalisations suivantes: le *tf* est normalisé par le maximum des *tf* de chaque doxel, et l'*idf* est basé sur le logarithme, selon la fréquence du terme dans la collection. Selon la pondération du système SMART, on est donc dans un schéma mt.

### 3.2. Modèle de mobilité

Le modèle de mobilité est limité dans cette première étude au contexte matériel de tâche utilisateur. Il est donné par  $ProfilMateriel = \langle M, P_M, R_{Mprofil} \rangle$  où :

- $M$  est l'ensemble des matériels possibles pour utiliser le système, tel que  $M = \{m\}$  avec  $m$  un matériel déterminé dans un environnement donné. Dans notre cas, un matériel est le modèle du système informatique sur lequel est réalisée l'interaction entre l'utilisateur et le système de recherche d'information, et l'environnement décrit en fait comment la liaison se fait avec le système de recherche d'information. Par exemple, un élément de  $M$  dénoterait un assistant personnel digital Toshiba e830 relié par un réseau Wifi de vitesse moyenne ;
- $P_M$  l'ensemble de tous les contextes matériels possibles ;
- $F_{Mprofil}$  : fonction de  $M$  vers  $P_M$  telle que  $F_{Mprofil}(m)=p$  représente le fait que le matériel  $m$  a un profil de matériel  $p$ .

Pour caractériser le profil matériel utilisé lors d'une recherche par un utilisateur, il est possible d'utiliser de nombreux paramètres, comme la vitesse du processeur, la taille de la mémoire centrale, la vitesse du bus entre processeur et mémoire, etc. Nous avons choisi dans cet article de limiter un profil de matériel aux capacités d'affichage des écrans et aux capacités de débit réseau. Un contexte matériel est décrit par  $P_M = \langle aff, net \rangle$  où :

- *aff* est l'intervalle  $[0, 1]$  et dénote la capacité d'affichage en terme de surface. Une valeur faible (proche de zéro) sera assignée à des téléphones portables par exemple (pour un Samsung E730 la résolution est de 176x220 pixels). Une valeur moyenne est assignée à des assistants digitaux personnels (pour un

PocketPC e830 de Toshiba la résolution est de 480 x 640 pixels) ou dans une moindre mesure à des consoles de jeux portables (comme la console PSP de Sony avec une résolution de 480 x 272 ou un iPhone d'Apple avec un écran de 320x480). Une valeur proche de 1 peut être assignée pour des ordinateurs de bureau ou portables (des résolutions de 1200 x 1600 ne sont pas rares pour de tels appareils).

- $net$  est l'intervalle  $[0, 1]$  et dénote la capacité du débit réseau utilisé entre le noyau fonctionnel du système de recherche d'information et l'interface de cette même application. Ce débit peut avoir une certaine influence sur les informations présentées à l'utilisateur : un faible débit peut nécessiter d'afficher peu d'informations à la fois, alors qu'un haut débit offre la possibilité d'acquérir davantage d'information en une fois. L'utilisation d'un modem téléphonique proposant un débit de 56Kb/s offre un débit faible, caractérisé par des valeurs proches de zéro. Des connections Bluetooth ou de l'adsl lent ( $\leq 1024$  kb/s) ont des valeurs moyennes. L'utilisation de réseaux filaires, Wifi g, ou bien de l'adsl rapide ( $> 2$  Mb/s) offre des débits élevés proches de 1.

Ces deux paramètres ne sont pas totalement corrélés, car on peut très bien utiliser un ordinateur portable avec un débit réseau faible, ou un assistant numérique portable avec un réseau rapide.

Pour un profil matériel  $m$  donné, on utilise les fonctions  $Aff(m)$  et  $Net(m)$  pour dénoter ses capacités d'affichage et de réseau.

### 3.3. Modèle d'interrogation

Nous posons que le matériel est connu et donné par  $m \in M$  lors de l'expression d'une requête d'un utilisateur.

Nous calculons  $RSV(d,q,m)$  la valeur de pertinence d'un doxel  $d$  pour une requête  $q$  dans un contexte de mobilité  $m$ . Le processus que nous proposons pour obtenir des résultats pertinents prenant en compte ce contexte en mobilité est décomposé en deux étapes : une étape de calcul de correspondance habituel en recherche d'information, puis une étape d'intégration de ces valeurs dans une formule qui utilise le contexte matériel de recherche.

Le premier calcul de pertinence des doxels  $d$  pour une requête  $q$  est effectué en utilisant le cosinus entre les vecteurs documents et le vecteur requête. Le résultat de cette étape est :

$$RSV_0(d,q) = \cos(\vec{d}, \vec{q})$$

Pour intégrer les aspects liés au matériel, nous proposons de mettre à jour les valeurs de pertinence  $RSV_0$  de la manière suivante :

- $Aff(m)$  est à prendre en compte pour indiquer que le document est moins pertinent en contexte matériel si le document n'est pas adapté au contexte

matériel pour l'affichage du contenu textuel du document. Par exemple, si l'affichage n'est pas grand et si le document est long, le document n'est pas adapté ;

- $Net(m)$  est utilisé pour indiquer si un document est adapté au matériel du point de vue de ses caractéristiques réseau. Par exemple, si le réseau est lent et si le document est long alors il est peu adapté.

Cette étape de prise en compte du matériel calcule les valeurs de pertinence en mobilité. On estime que la prise en compte des paramètres d'affichage et de débit réseau est plutôt multiplicatif sur la valeur de pertinence finale, car chacun de ces éléments peut être la cause d'une faible valeur globale de pertinence. Il en résulte que la correspondance finale  $RSV(d,q,m)$  est telle que :

$$RSV(d,q,m) = RSV_0(d, q) * f_{CM}(d, m)$$

Avec  $f_{CM}$  la fonction qui renvoie une valeur dans  $[0,1]$  en prenant en compte le document  $d$  et le contexte matériel  $m$ .

Ci-dessous, nous proposons un exemple de fonction  $f_{HC}$  qui lie les paramètres de contexte matériel avec la taille (en octets) du contenu textuel des doxels considérés (indiqué par fonction  $taille$ ) :

$$f_{CM}(d, m) = Aff(F_{Mprofil}(m)) * (1/taille(d)) * Net(F_{Mprofil}(m)) * (1/taille(d))$$

## 4. Evaluation de systèmes de recherche de documents structurés en contexte de mobilité

Rappelons que nous choisissons une approche inspirée de la méthode du gain cumulé, car nous voulons prendre en compte le fait que tous les documents n'ont pas la même valeur de pertinence. Dans notre cas spécifique, nous voulons être en mesure d'exprimer que la taille des documents renvoyés influe sur la valeur de pertinence des documents.

De manière plus détaillée, nous considérons que cette valeur de pertinence doit intégrer les éléments suivants pour un doxel  $d$  pertinent :

- sa taille, car nous tentons de privilégier les documents pertinents les plus petits renvoyés,
- sa proportion de pertinence basée sur le taux de recouvrement entre le doxel  $d$  et le ou les doxels de la liste des pertinents. Ceci détermine la proportion de pertinence du doxel renvoyé. Plus cette proportion est grande, mieux c'est.
- la pertinence du ou des éléments de la liste des pertinents qui correspondent à  $d$ , prenant en compte les parties pertinentes les plus petites et les plus grandes de la liste des pertinents. Ce paramètre se base sur les



caractéristiques globales de la liste des pertinents, et permet de contrebalancer l'importance des positions si tous les doxels pertinents ont une taille similaire.

Nous proposons donc une mesure qui prend en compte les tailles des documents pertinents ainsi que la taille globale des résultats renvoyés par le système. Comme il existe plusieurs cas d'intersection entre un doxel renvoyé et un doxel pertinent, nous devons prendre en compte ces différentes possibilités, tout en s'assurant qu'à la limite ces fonctions utilisées sont continues :

- Nous utilisons une liste idéale de doxels pertinents les plus petits, triée par la taille de ces documents décroissante :  $LI[i]$ . Cette liste est similaire à ce qui a été décrit en partie 3, et ne possède aucun recouvrement structurel, ce qui veut dire qu'il n'existe aucune relation de composition (directe ou non) entre deux éléments de  $LI$ . Nous appelons  $|LI|$  le nombre de doxels de  $LI$ . Pour chaque doxel  $d_i$  de  $LI$ , nous connaissons la taille de sa partie pertinente en nombre de caractères par la fonction  $size_p(d_i)$ . Par construction, la taille maximale des doxels de  $LI$  est  $size_p(LI[1])$ , et la taille minimale est  $size_p(LI[|LI|])$ . Pour un doxel  $d$ , nous posons qu'il existe une fonction  $pos(LI, d)$  qui renvoie la position de  $d$  dans  $LI$  si  $d$  est un doxel pertinent. Pour un doxel  $d$ , nous définissons une fonction  $pos\_composing(LI, d)$  qui renvoie l'ensemble des doxels pertinents de  $LI$  qui composent  $d$ . Pour un doxel  $d$ , nous définissons une fonction  $pos\_composed(LI, d)$  qui donne le doxel de  $LI$  qui est composé de  $d$ .
- Nous obtenons la liste des documents retournés par le système de recherche d'information testé :  $LR[j]$ , contenant  $|LR|$  éléments.
- La valeur de pertinence de gain pour un élément  $d (=LR[j])$  est la valeur de gain en contexte  $xG_c[j]$  définie par :
  - o Si le doxel  $d$  est élément de  $LI$ , sa valeur de gain est fonction de la position de  $d$  dans  $LI$  (première partie de la formule [4]), de la proportion de  $d$  qui est pertinente pour la requête (deuxième partie de la formule [1]), et fonction de la taille de la partie pertinente par rapport à la collection des pertinents  $LI$ . Dans la formule [4], la première partie est une expression qui est linéairement décroissante en fonction de la position. Une telle expression permet de ne pas avoir un impact prépondérant de cette position dans la formule, ce qui arriverait si on utilisait une simple expression comme  $1/pos(LI, d)$  par exemple. La dernière partie de la formule [4] effectue un changement d'échelle qui consiste à calculer la différence entre la taille minimale des pertinents et la taille du doxel  $d$ , normalisée par la différence entre la taille de doxel la plus grande moins la taille de doxel la plus petite ; cette valeur est donc dans l'intervalle  $[0, 1]$ , puis ôte cette valeur de 1 afin de favoriser les documents les plus petits.

$$xG_c[j] = \left(1 - \frac{\text{pos}(LI, d) - 1}{|LI|}\right) \times \frac{\text{size}_p(d)}{\text{size}(d)} \quad [4]$$

$$\times \left(1 - \frac{\text{size}_p(d) - \text{size}_p(LI[1])}{\text{size}_p(LI[|LI|]) - \text{size}_p(LI[1])}\right)$$

Soit un doxel  $d$ , de taille 60 caractères avec 30 caractères pertinents, fourni en réponse par un système de recherche d'information. Supposons que  $d$  est en position 50 dans la liste  $LI$ . Supposons également que la liste  $LI$  possède 300 doxels, que le doxel de  $LI$  le plus petit a une taille de 40 et que le doxel le plus grand de  $LI$  a une taille de 1000. Alors la valeur de gain  $xG_x[d]$  est égale à

$$(1 - 50/300) * 30/60 * (1 - 260/960) = 0.3038$$

- Si  $d$  est englobé par un élément  $d'$  de  $LI^3$ , alors nous devons modifier la formule [4] en intégrant le fait que l'on ne connaît pas le réel ratio de pertinence de  $d$ . Nous l'évaluons alors en considérant que la pertinence de  $d$  est le produit du ratio de pertinence de  $d'$  et de la taille relative de  $d$  par rapport à  $d'$ . Nous obtenons donc dans ce cas la formule [5]. Dans le cas limite où  $d'$  est uniquement composé de  $d$ , alors nous obtenons la formule [4] car  $\text{size}(d') = \text{size}(d)$ , il y a donc bien continuité entre ces deux formules.

$$xG_c[j] = \left(1 - \frac{\text{pos}(LI, d') - 1}{|LI|}\right) \times \frac{\text{size}_p(d')}{\text{size}(d')} \times \frac{\text{size}(d)}{\text{size}(d')} \quad [5]$$

$$\times \left(1 - \frac{\text{size}_p(d') - \text{size}_p(LI[1])}{\text{size}_p(LI[|LI|]) - \text{size}_p(LI[1])}\right)$$

Soit un doxel  $d$ , de taille 60 caractères, fourni en réponse par un système de recherche d'information. Supposons que  $d$  est contenu par le doxels  $d'$  en position 50 dans la liste  $LI$ , la taille de  $d'$  étant de 100 caractères avec 75 caractères pertinents. Supposons également que la liste  $LI$  possède 300 doxels, que le doxel de  $LI$  le plus petit a une taille de 40 et que le doxel le plus grand de  $LI$  a une taille de 1000. Alors la valeur de gain de  $d$  est égale à :

$$(1 - 50/300) * 75/100 * 60/100 * (1 - 35/960) = 0.3613$$

- Si  $d$  englobe un élément ou plusieurs éléments  $D' = \{d'\}$  de  $LI^4$ , nous intégrons ce fait en nous inspirant de la formule [4], et en sommant la valeur de pertinence pour chaque élément  $d'$ . Dans le cas limite où  $d$

<sup>3</sup> Par hypothèse sur  $LI$  il n'y a qu'un élément de  $LI$  au maximum qui englobe  $d$ .

<sup>4</sup> Par hypothèse sur  $LI$  il peut y avoir plusieurs éléments de  $LI$  qui sont englobés par  $d$ .

est en fait composé uniquement d'un  $d'$ , la formule [6] est la même que la formule [4].

$$xG_c[j] = \sum_{d' \in \text{pos\_composin } g(LI, d)} \left(1 - \frac{\text{pos}(LI, d') - 1}{|LI|}\right) \times \frac{\text{size}_p(d')}{\text{size}(d)} \times \left(1 - \frac{\text{size}_p(d') - \text{size}_p(LI[1])}{\text{size}_p(LI[|LI|]) - \text{size}_p(LI[1])}\right) \quad [6]$$

Soit un doxel  $d$ , de taille 400 caractères, fourni en réponse par un système de recherche d'information. Supposons que  $d$  contient deux doxels :  $d'$  de taille 100 caractères avec 75 caractères pertinents en position 50 dans la liste  $LI$ ,  $d''$  de taille 180 caractères avec 80 caractères pertinents en position 60 dans la liste  $LI$ . Supposons également que la liste  $LI$  possède 300 doxels, que le doxel de  $LI$  le plus petit a une taille de 40 et que le doxel le plus grand de  $LI$  a une taille de 1000. Alors la valeur de gain de  $d$  est égale à :

$$(1 - 50/300) * 75/400 * (1 - 35/960) + (1 - 60/300) * 80/400 * (1 - 40/960) = 0.3039$$

Les trois cas ci-dessus sont exclusifs, donc il n'y a pas d'ambiguïté sur l'usage de l'une ou l'autre des formules.

A partir de ces valeurs de gain, on définit une formule [7] similaire à la formule [1] pour le gain cumulé en contexte :

$$xCG_c[i] = \sum_{j=1}^i xG_c[j] \quad [7]$$

Ensuite, pour calculer le gain en contexte normalisé, on utilise les formules [2] et [3] de la même manière que défini par INEX 2006. Nous notons  $nxCG_c[i]$  la valeur de gain cumulé jusqu'au  $i^{\text{ème}}$  élément de la liste renvoyée par un système.

Pour obtenir une valeur de gain cumulé en contexte sur plusieurs requêtes, nous faisons la moyenne des  $xCG_c[i]$  pour l'ensemble des requêtes considérées.

## 5. Résultats

Nous évaluons notre proposition de recherche de documents structurés en contexte grâce aux valeurs de gain cumulé utilisées par INEX 2006 et avec notre extension du gain cumulé. Pour cela, nous nous limitons dans cet article à des valeurs de gain cumulé en contexte à 5 et 10 doxels. L'objectif de cette évaluation est de montrer que nos mesures d'évaluations sont davantage adaptées dans le cas de la recherche

d'information en mobilité favorisant le renvoi par un système de petits documents. Le corpus d'évaluation est celui de 2006 (600 000 documents), et nous avons utilisé les 7 requêtes suivantes : 289, 290, 293, 294, 295, 296, 297.

Dans la table 1, nous montrons les résultats obtenus dans une recherche « hors contexte », notée *HC*, et avec une recherche en contexte noté *EC<sub>0.4,0.4</sub>*, en fixant les paramètres de contexte matériel de mobilité ayant une capacité d'affichage de *0.4*, et une capacité de réseau de *0.4*. La recherche hors contexte utilise la fonction *RSV<sub>0</sub>* et la recherche en contexte la fonction *RSV*, définies en partie 3.3. Lors des tests en contexte, la fonction utilisée pour estimer la taille des doxels est la norme des vecteurs les décrivant. Nous évaluons les résultats avec le calcul de gain cumulé normalisé de INEX 2005, appelé *nxCG<sub>inex</sub>*, et notre évaluation en contexte, appelée *nxCG<sub>c</sub>*.

Pour les résultats en gain cumulé *nxCG<sub>inex</sub>* la configuration en contexte fournit des meilleurs résultats, avec +179% et +313% par rapport au défaut considéré. Ceci veut dire que pour les mesures de gain cumulé le système avec utilisation du contexte s'améliore au fur et à mesure des résultats. Par contre, les résultats obtenus avec le gain cumulé en contexte donne des meilleurs résultats (relativement) après 5 documents qu'après 10 avec respectivement +53 et +23%. On conclut des résultats présentés dans ce tableau deux éléments clés :

- notre mesure a bien pour résultat de favoriser les systèmes qui prennent en compte explicitement de la taille des documents dans leurs calculs,
- le système en contexte fournit des résultats pertinents et petits dans les tous premiers documents ce qui était attendu. Cependant, si les résultats obtenus par le système en contexte restent relativement stables en terme de gain cumulé en contexte (passage de 1,3 à 1,7 E-3), le système hors contexte lui semble s'améliorer entre les documents 5 et 10 en moyenne. Ce point important n'est absolument pas détecté par les mesures originelles de gain cumulé d'INEX. Les valeurs de gain cumulé normalisé d'INEX indiquent qu'en fait le système en contexte renvoie bien des documents pertinents, sans vraiment fournir les plus petits, mais il obtient cependant de meilleurs résultats que l'approche qui n'utilise pas les tailles des documents.

	<i>nxCG<sub>inex</sub></i> [5]	<i>nxCG<sub>inex</sub></i> [10]	<i>nxCG<sub>c</sub></i> [5]	<i>nxCG<sub>c</sub></i> [10]
HC	0,1843	0,1071	$7.25^E-4$	$1.27^E-3$
<i>EC<sub>0.4,0.4</sub></i>	0,5143 (+ 179%)	0,4429 (+ 313%)	$1.33^E-3$ (+ 83 %)	$1.74^E-3$ (+ 36 %)

**Table 1. Calculs de *nxCG<sub>c</sub>*.**

## 6. Conclusion

Nous avons proposé dans cet article deux éléments : le premier a trait à la définition d'un modèle relativement simple de recherche de documents structurés en situation de mobilité dans lequel on tente de renvoyer des documents tenant compte du fait

que l'interface de recherche et de visualisation des résultats est petite, le second visait à proposer une méthode pour évaluer de tels systèmes en se basant sur les approches à base de gain cumulé. Nous avons choisi d'étudier à la fois une approche de recherche de petites parties de documents pertinentes et une mesure d'évaluation afférente, ce qui peut paraître paradoxal. L'approche simple de prise en compte de contexte matériel utilisée a en fait pour objectif principal de vérifier que la mesure que nous proposons est viable.

Les résultats que nous avons obtenus sont conformes à nos attentes : quand le système que nous utilisons tient compte du contexte matériel en mobilité, alors il obtient de meilleurs résultats qu'un système ne tenant pas compte de ce contexte. Il faut souligner que notre évaluation ne s'applique pas uniquement à notre système, mais vise un cadre plus large décrivant des applications de recherche de documents en situation mobile avec des faibles capacités d'affichage et d'interaction.

Ce travail est un premier pas pour permettre une évaluation de systèmes en contexte de mobilité. Ce genre d'évaluation n'existe pas et devra être développé pour comparer les systèmes futurs de recherche d'information.

Ce travail est préliminaire, nous allons expérimenter sur les corpus INEX 2006 et 2007 complets, pour déterminer sa robustesse face à de plus grandes variabilités que ce qui a été décrit ici. De plus, nous évaluerons plus en détail le comportement de notre mesure de gain par rapport aux autres mesures existant sur les documents structurés afin de montrer son intérêt.

Nous allons également par ailleurs développer notre modèle de recherche de documents structurés pour prendre en compte davantage les aspects de mobilité.

## Bibliographie

- F. Crestani, H. Du, Written versus spoken queries: A qualitative and quantitative comparative analysis, *Journal of the American Society for Information Science and Technology*, Volume 57, Issue 7, pp. 881-890, 2006.
- F. Crestani, H. Du., Written versus spoken queries:: An Experiment on Mandarin Chinese. *IJCNLP 2004*, pp. 745-754, 2004.
- P. Ingwersen, K. Järvelin, Information Retrieval in Contest – Irix, *ACM SIGIR Forum* 39/2 December 2005.
- D. H. Fang Huang, S. Watt and M. Clark. Robert Gordon University at INEX 2006: Adhoc Track. In *INEX 2006 Workshop Pre-Proceeding*, pages 70-79, 2006.
- M. Lalmas, G; Kazaï, J. Kamps, J. Pehcevski, B. Piwowarski, S. Roberston, INEX 2006, Evaluation Measures, 5th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2006, Dagstuhl Castle, Germany, December 17-20, 2006.
- N. Milic-Frayling, R. Sommerer, K. Rodden, and A. F. Blackwell, Searchmobil: Web viewing and search for mobile devices. In *WWW (Posters)*, 2003.
- X. Yin , W. S. Lee, Using link analysis to improve layout on mobile devices, *Proceedings of the 13th international conference on World Wide Web*, pp. 338-344, 2004.