

Évaluation des Interfaces Utilisateur d'Information

Nicolas Bonnel*, Max Chevalier****

*IRISA / Université de Rennes 1, Campus universitaire de Baulieu, 35042 Rennes Cedex
nicolas.bonnel@irisa.fr
<http://www.irisa.fr/texmex/Nicolas.Bonnel>

**IRIT / Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex
Max.Chevalier@irit.fr
<http://www.irit.fr/~Max.Chevalier>

***LGC / IUT Rangueil, 115 route de Narbonne, 31062 Toulouse Cedex

Résumé. Tout processus de recherche d'information (RI) n'a de sens aux yeux des utilisateurs qu'à travers l'ultime étape qui consiste à visualiser les résultats. L'importance que prend la visualisation (ou restitution) des résultats est à l'origine des nombreuses propositions d'interfaces, qu'elles soient textuelles, 2D ou 3D. Si des évaluations de certaines de ces interfaces ont été proposées, aucune comparaison n'a réellement été réalisée dans ce contexte faute de contraintes sur ces interfaces et de critères de comparaison relatifs à la tâche de RI. Dans cet article, nous proposons d'introduire des pistes d'évaluation afin d'aboutir à un cadre expérimental permettant l'évaluation et ainsi la comparaison des *Interfaces Utilisateur d'Information*.

Mots-clés. Interface Utilisateur d'Information, interface de restitution de résultats de recherche, systèmes de recherche d'information, critères d'évaluation et de comparaison, évaluation « en contexte ».

1 Introduction

Pour rechercher des informations sur le Web, une grande majorité des utilisateurs utilise des moteurs de recherche. Ces outils proposent un frontal grâce auquel l'utilisateur peut spécifier sa requête et en visualiser les résultats. On constate alors un décalage entre le nom de ces outils et leur fonction du point de vue des utilisateurs. C'est pourquoi nous parlerons désormais dans cet article d'**Interfaces Utilisateur d'Information (ou IUI)**. Ce terme a été proposé par Franck Poisson, ex-Directeur Général et fondateur du bureau France de GOOGLE.

L'évaluation des IUI est une tâche particulièrement difficile. Jusqu'ici, la méthode la plus courante est la réalisation d'une étude (ou test) utilisateur. Celle-ci ne nous semble pas suffisante et nous proposons dans les sections suivantes des pistes d'évaluation des IUI. Nous présentons tout d'abord les principales limites des évaluations faites actuellement qui sont

certes très instructives mais qui ne nous permettent pas d'évaluer l'impact qu'a l'IUI sur la tâche accomplie par l'utilisateur. Nous proposons ensuite un ensemble de critères nous permettant d'évaluer, de comparer et de classer les différentes IUI en fonction des tâches, des utilisateurs...

Notre approche est présentée dans le cadre spécifique des interfaces de visualisation de résultats de recherche. Ces résultats peuvent provenir de moteurs de recherche (cadre applicatif privilégié dans les exemples de la section 2) ou plus généralement de systèmes de recherche d'information. Cependant notre approche reste adaptée à d'autres cadres applicatifs similaires tels que la visualisation exploratoire de données ou la fouille visuelle de données.

2 Exemples d'Interfaces Utilisateur d'Information

De nombreuses IUI ont déjà été proposées. Certaines sont utilisées alors que d'autres n'ont pas dépassées le stade du prototype. Cependant le problème majeur reste les trop grandes disparités entre ces différentes interfaces, ce qui rend toute comparaison impossible. Afin d'illustrer notre propos, nous citons ci-après quelques exemples d'IUI majoritairement proposées dans le cadre de la RI sur le Web. La plus classique et la plus utilisée est celle proposée par GOOGLE¹. Il s'agit d'une interface (**voir figure 1**) qui propose un affichage linéaire des résultats de recherche sous forme d'une liste triée selon un critère de « pertinence ». VIVÍSIMO² reprend ce principe mais en y ajoutant une catégorisation automatique des résultats dans une hiérarchie de répertoires significatifs, *via* une technique de *clustering* à la volée. L'interface obtenue (**voir figure 2**) est alors un ensemble de listes toujours triées selon un critère de « pertinence ». L'interface UJIKO³ est un autre exemple de catégorisation automatique des résultats (**voir figure 3**). Cet aspect classification des résultats est aussi repris dans des interfaces 2D, telles que GROKKER⁴ qui organise les résultats dans des catégories (**voir figure 4**). D'autres interfaces proposent des visualisations sous forme cartographique, généralement basées sur les relations qui existent entre les différents résultats. C'est le cas du méta-moteur de recherche KARTOO⁵ qui présente les résultats sous forme cartographique (**voir figure 5**). Dans cette interface l'utilisateur peut visualiser les liens entre les résultats et certains « mots-clés ». Cependant l'aspect linéaire persiste étant donné que les résultats sont présentés sous la forme d'une succession de cartes. D'autres interfaces vont plus loin en exploitant la possible spatialisation du Web. Cela se retrouve notamment dans des IUI basées sur l'affichage des résultats dans un environnement virtuel en 3D, telles que VIOS⁶ (**voir figure 6**). Ce concept est aussi utilisé dans le prototype SMARTWEB (Bonnell et al., 2005a,b) qui représente les résultats dans une ville virtuelle en 3D (**voir figure 7**). Mais, contrairement à VIOS, SMARTWEB organise les résultats dans cet espace 3D grâce au calcul d'une carte auto-organisatrice, ce qui permet de regrouper (et de placer) les résultats selon la distribution des mots et ainsi d'avoir une « proximité sémantique ». Il existe également d'autres approches, notamment avec des visualisations 3D.

¹ <http://www.google.com>

² <http://www.vivisimo.com>

³ <http://www.ujiko.com>

⁴ <http://www.grokker.com>

⁵ <http://www.kartoo.com>

⁶ *Visual Internet Operating System*, <http://computer.howstuffworks.com/vios.htm>

Par exemple, CAT-A-CONE (Hearst et Karadi, 1997) utilise une visualisation 3D d'un arbre (*Cone Trees* (Robertson et al., 1991)) afin d'afficher simultanément les résultats obtenus et une hiérarchie de catégories prédéfinies (voir **figure 8**). Un autre exemple est le prototype NIRVE (Cugini et al., 2000) qui permet à l'utilisateur de visualiser et manipuler un ensemble de documents résultant d'une requête sur un moteur de recherche. Il possède plusieurs méthodes de visualisation dont un exemple est proposé sur la **figure 9**. Dans cet exemple, appelé « *Spoke and Wheel Model* », l'utilisateur peut agréger des mots-clés afin de former un plus petit ensemble de concepts. À partir de ces concepts, des groupes de documents sont créés et affichés de telle façon que la distance angulaire entre les groupes soit proportionnelle à la distance logique entre eux. De même les distances radiales entre les documents reflètent les distances métriques qui séparent leurs profils (basés sur les concepts définis par l'utilisateur). On peut également citer d'autres interfaces telles que EASY-DOR (Chevalier, 2002) qui utilisent un affichage 3D basé sur des axes représentant chacun un ou plusieurs mots-clés de la requête (voir **figure 10**).



FIG. 1 – GOOGLE.

Évaluation des Interfaces Utilisateur d'Information

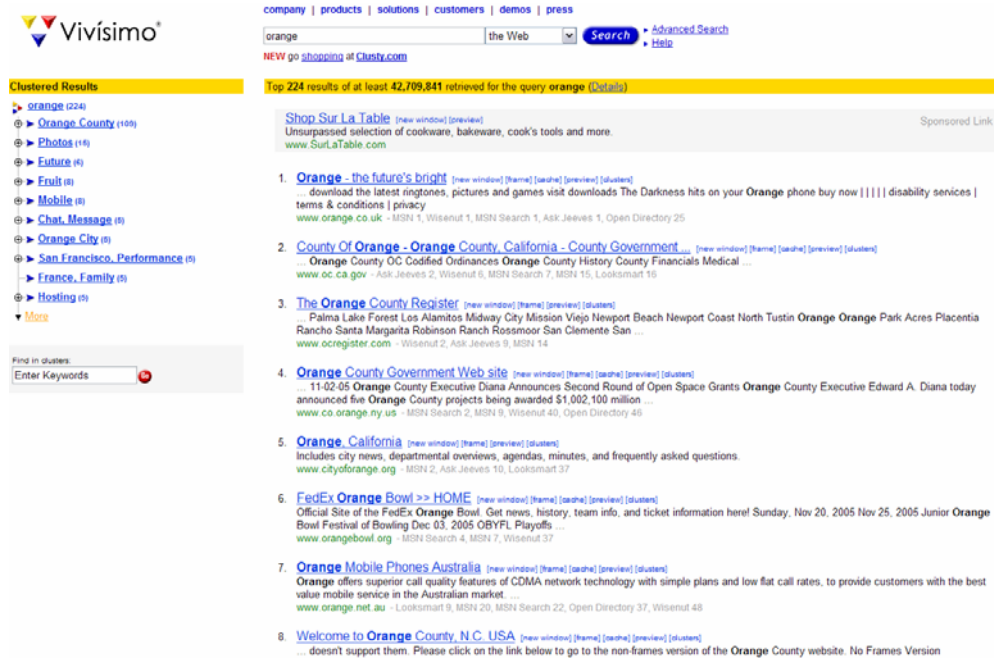


FIG. 2 – Vivísimo.

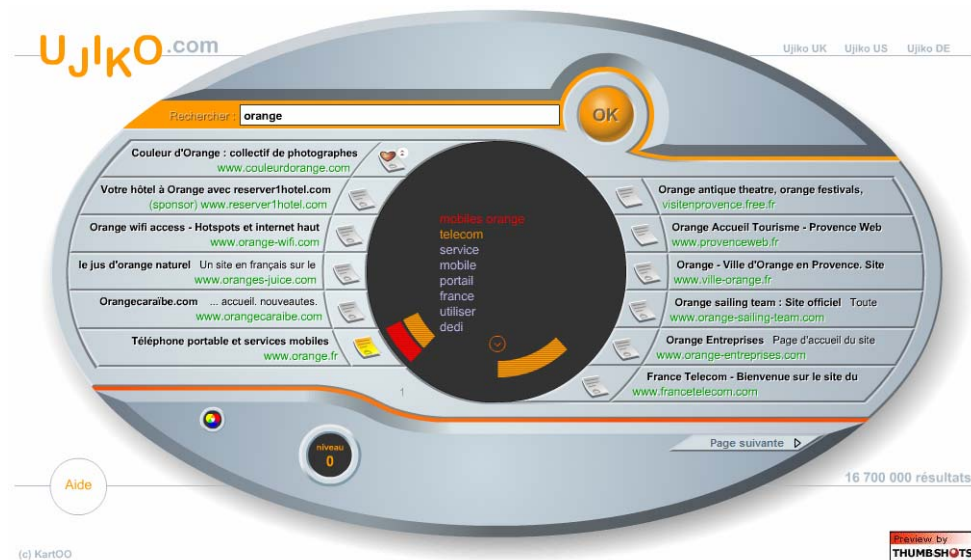


FIG. 3 – UJKO.

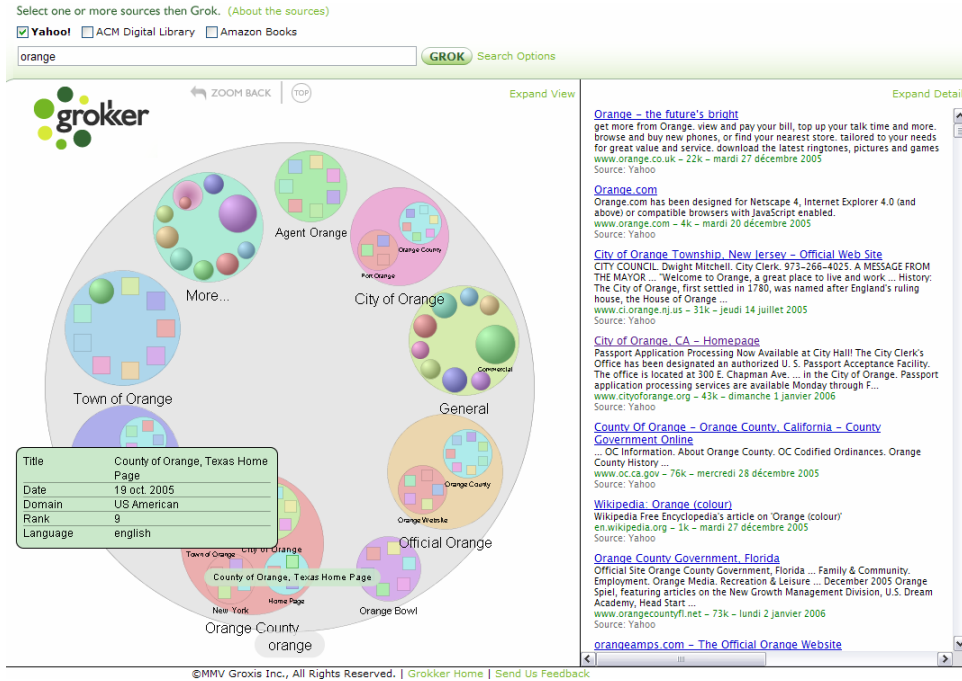


FIG. 4 – GROKKER.



FIG. 5 – KARTOO.

Évaluation des Interfaces Utilisateur d'Information



FIG. 6 – VIOS.



FIG. 7 – SMARTWEB.

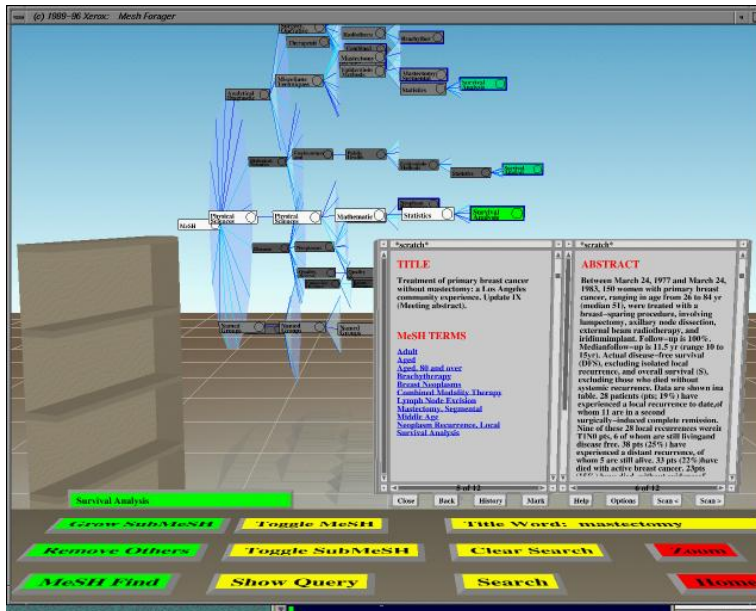


FIG. 8 – CAT-A-CONE.

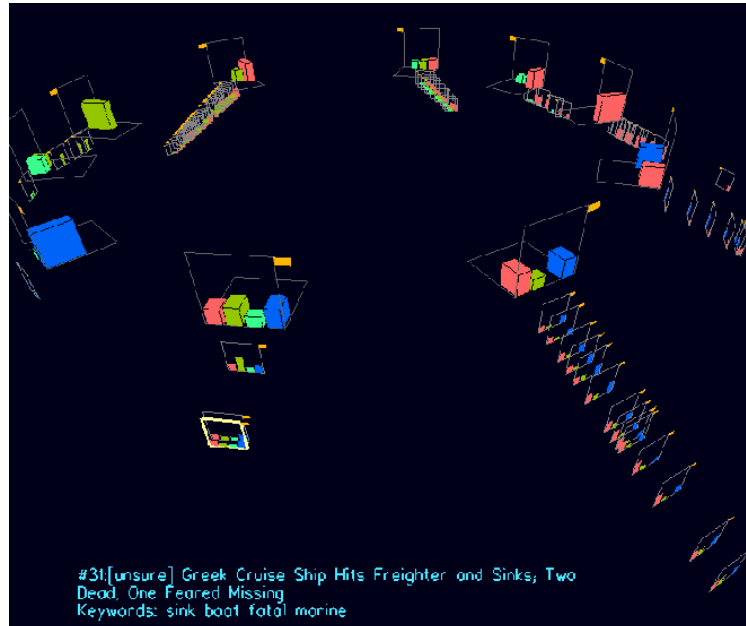


FIG. 9 – NIRVE, Spoke and Wheel Model.

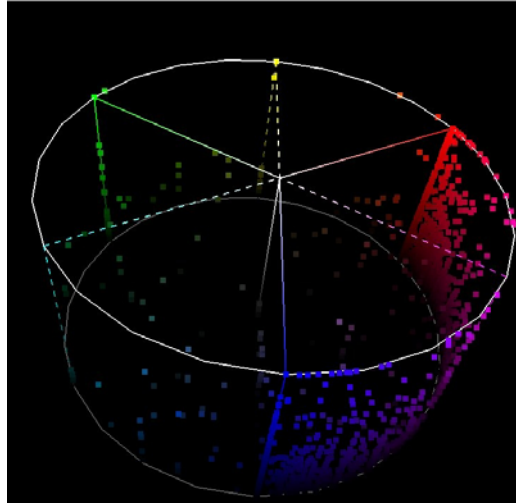


FIG. 10 – *EASY-DOR*.

Toutes ces IUI ne reposent pas sur le même système de recherche et ne proposent pas les mêmes traitements (tels que le *clustering* ou le filtrage des résultats). Aussi, bien que des évaluations isolées de ces IUI soient possibles, elles sont généralement peu interprétables et il est impossible de les comparer de façon efficace, surtout au regard de la tâche de RI. Nous proposons donc dans la suite de cet article de s'intéresser aux conditions nécessaires pour la mise en place d'un cadre expérimental d'évaluation et de comparaison de ces IUI. Nous distinguons alors les contraintes (**section 4**) qui doivent être prises en compte lors de la phase d'évaluation (mais aussi lors de la conception de ces interfaces), et les critères d'évaluation (**section 5**) qui serviront à l'évaluation et à la comparaison avec d'autres systèmes de visualisation. Mais, dans un premier temps, nous nous intéressons au cadre général d'évaluation d'une IUI (**section 3**).

3 Évaluation d'une IUI

La méthode la plus courante pour évaluer une IUI est la réalisation d'une étude utilisateur. Cette technique consiste à proposer un questionnaire aux utilisateurs afin qu'ils répondent à chaque question généralement par une note. Ce questionnaire est très souvent inspiré des propositions de Shneiderman (1998) en matière d'évaluation d'interfaces utilisateur et est très orienté « utilisabilité ». Ce principe a notamment été adopté dans (Chevalier, 2002) et (Bonnell et al., 2005c). Cependant, il s'avère que les résultats de ces tests, dans le cas de l'évaluation d'une IUI, sont difficilement interprétables. En effet, les questions posées ne sont pas toujours adaptées au contexte des IUI et il n'est pas possible de vraiment apprécier l'impact de l'IUI sur les performances (du point de vue de l'utilisateur) du processus de recherche d'information (Chevalier et Hubert, 2005).

Dans le cadre de l'évaluation des IUI, deux approches peuvent être distinguées : l'évaluation d'une IUI utilisant son propre système de recherche (*i.e.* ses propres résultats de

recherche) et l'évaluation d'une IUI indépendante de tout système de recherche. Dans le premier cas, nous pouvons retrouver les principaux *benchmarks* qui ont été mis en place dans le contexte de la recherche d'information, et notamment ceux rattachés aux campagnes d'évaluation TREC⁷, CLEF⁸ ou encore INEX⁹. Or ces campagnes d'évaluation n'étaient pas initialement prévues pour évaluer l'interface utilisateur, ce qui explique l'absence de certains critères spécifiques aux IUI. De plus, dans ces campagnes d'évaluation, les IUI sont évaluées en même temps que le système de recherche propre à chaque participant. Ainsi, même si l'interaction est mise en évidence, l'impact de l'IUI est noyé dans les performances intrinsèques du système de recherche. Cette façon d'évaluer ne permet pas, en soi, de définir si une IUI est performante ou non pour les tâches auxquelles elle est destinée. Ce type d'évaluation suppose que l'intérêt de l'IUI est corrélé à la performance du système. La question qu'il faudrait alors se poser est : comment se comporterait une bonne interface avec un système de recherche peu performant ? Cette question peut faire sourire et nous pensons tous avoir la réponse. Or, nous pouvons penser que l'interface peut « relever » les performances du système. Par exemple, une IUI basée sur une classification des résultats peut proposer des classes tout à fait pertinentes même si le système ne fait pas de la haute précision. Par ailleurs, une comparaison avec d'autres systèmes de recherche reste « possible » mais elle ne permet pas d'identifier les raisons d'un succès ou d'un échec de l'IUI (est-ce lié au système de recherche ou à l'interface ?). Outre ces campagnes d'évaluation orientées RI, nous pouvons souligner l'existence de campagnes d'évaluation plus orientées interaction homme-machine telles que (Fekete, 2004), qui fournissent une vision complémentaire de l'évaluation d'une IUI « en contexte ». Nous regrettons cependant que ces différentes campagnes d'évaluation n'aient pas plus « d'interactions » entre elles.

Dans cet article, nous nous intéressons donc essentiellement au deuxième cas, c'est-à-dire celui où l'évaluation des IUI peut se faire indépendamment du système de recherche. Si jusqu'à présent les systèmes « les plus chanceux » avaient le privilège d'être évalués, les comparaisons étaient quasi-inexistantes du fait de la trop grande hétérogénéité des systèmes. Dans un premier temps, ce n'est pas tant la comparaison intrinsèque qui nous intéresse mais plus l'évaluation des IUI en prévision de les classifier afin d'identifier, selon le contexte d'utilisation, quelles IUI sont les plus adéquates. Cependant, le cadre d'évaluation de certains critères proposés dans la section 5, peut rendre possible la comparaison de différentes IUI. Par ailleurs, nous sommes convaincus que pour qu'un système de recherche rende les meilleurs services aux usagers, il est nécessaire qu'il propose différentes IUI. Le choix de l'IUI peut alors être, soit proposé explicitement à l'utilisateur qui choisit en fonction de la tâche à réaliser, soit pris en compte dans l'adaptabilité automatique ou semi-automatique du système par rapport aux données à visualiser (ou aussi par rapport à sa connaissance du profil de l'utilisateur). Plusieurs IUI pour un système de recherche permettent alors de répondre aux besoins aussi variés que le sont les usagers. En effet, Shneiderman (1998) soulignait qu'une interface peut ne pas répondre à toutes les tâches et à tous les besoins des usagers. Nous pouvons ainsi faire apparaître que l'utilité d'une IUI dépend donc, en partie, de différents critères liés à l'utilisateur. Pour catégoriser ces critères nous pouvons reprendre la vision de Lainé-Cruzet (1999) qui souligne que l'utilisateur peut être caractérisé par trois questions : Qui

⁷ Text REtrieval Conference - <http://trec.nist.gov>

⁸ Cross-Language Evaluation Forum - <http://clef.isti.cnr.it>

⁹ Initiative for the Evaluation of XML Retrieval - <http://inex.is.informatik.uni-duisburg.de/2005/>

est-il ? Que veut-il ? Pour quoi faire ? Ce ne sont pas les seuls critères qui doivent entrer en compte dans l'évaluation. En effet, il y a certains critères liés aux documents retrouvés en résultat. Ainsi, la visualisation en liste de résultats (*e.g.* GOOGLE) peut suffire pour une tâche donnée et/ou pour un nombre très restreint de documents. Avant de présenter les critères que nous avons identifiés, nous nous intéressons aux différentes contraintes que doit respecter, à notre avis, la phase d'évaluation.

4 La phase d'évaluation

Pour obtenir des résultats exploitables, il est nécessaire de préparer avec soin la phase d'évaluation. Notamment, afin de pouvoir évaluer et comparer efficacement différentes IUI, un certain nombre de contraintes doivent être respectées lors de la phase d'évaluation. Ces contraintes doivent être prises en compte dès la conception de l'interface afin de pouvoir garantir leur respect. Ainsi l'IUI doit être :

- **indépendante** du système de recherche (*i.e.* des données). Il est en effet nécessaire que l'IUI accède aux données via des interfaces proposées par les systèmes de recherche et libres d'utilisation ;
- **modulable** en terme de fonctionnalités, c'est-à-dire que les fonctionnalités doivent être adaptables au système de recherche utilisé. En effet, l'interface peut être destinée à plusieurs outils ou applications ;
- **capable** de gérer des jeux de requêtes prédéfinis.

Ces différents aspects sont présentés dans les sous-sections suivantes.

4.1 Données

Le premier point à vérifier est de s'assurer que les interfaces traitent le même ensemble de résultats, c'est-à-dire que les données visualisées sont identiques. Dans notre cas, les résultats peuvent être déclarés identiques à condition de respecter trois égalités : celle du contenu, celle du nombre et celle des descripteurs disponibles (*i.e.* les mêmes informations de base sur les documents sont disponibles mais chaque IUI peut les traiter différemment). La solution la plus simple pour répondre à cette contrainte fondamentale à l'évaluation et à la comparaison de deux IUI, est d'utiliser les API¹⁰ fournies par certains moteurs de recherche. Un système de recherche de référence pourrait être proposé si un moteur de recherche proposait une API suffisamment riche en termes de fonctionnalités. Malheureusement, les API existantes ne donnent pas accès à l'ensemble des descripteurs disponibles dans les bases d'index. Une solution pourrait alors être la proposition d'un système de recherche destiné à l'évaluation des IUI et fournissant une API riche. Il faudrait que l'API propose des fonctions d'accès suffisamment riches pour que chaque IUI y trouve son compte. Cependant, des besoins supplémentaires de la part d'une IUI peuvent exister, ce qui se traduirait par une différence en ressources nécessaires. Dans ce cas, il faudrait aussi que l'API soit suffisamment ouverte pour la faire évoluer et y intégrer ces nouveaux besoins afin d'améliorer le cadre d'évaluation et de comparaison de ces interfaces.

¹⁰ Application Programming Interface

4.2 Fonctionnalités

Il est nécessaire que les différentes IUI évaluées puissent s'adapter en fonction de l'API proposée ; les fonctionnalités de ces IUI étant très souvent dépendantes des informations fournies par l'API. Ainsi, en supposant que les moteurs de recherche adoptent une API partagée, l'utilisation d'une IUI avec n'importe quel moteur sera désormais possible. Ce n'est pas le cas aujourd'hui est cette vision reste à l'heure actuelle utopique.

4.3 Jeux de requêtes

Il est important que chaque IUI exploite les mêmes jeux de requêtes permettant de couvrir les différentes tâches que l'on souhaite évaluer. Par exemple, en recherche d'information, nous pourrions identifier les tâches suivantes :

- recherche d'un élément précis (exemple : existe-t-il une page web professionnelle de François Poulet ?). En général le résultat peut être assimilé à une information booléenne dans le sens où l'usager cherche une information très précise ;
- recherche de l'existence d'une information *a priori* connue (exemple : je recherche les sites web parlant des lois de Murphy) ;
- recherche exploratoire, c'est-à-dire panorama d'un thème donné (exemple : je recherche tout ce qu'il existe dans le domaine de la gestion des connaissances).

Chacune de ces tâches n'implique pas, par leur nature, la même restitution en termes d'IUI. Les tâches doivent donc être définies avant la phase d'évaluation car les jeux de requêtes devront répondre aux différentes tâches identifiées que l'on souhaite évaluer. De plus, chacune des tâches identifiées devra être caractérisée par un ou plusieurs critères (définis en section 5) qui pourront être éventuellement pondérés. Par exemple, nous pouvons imaginer que la tâche « recherche d'un document précis » soit caractérisée entre autres par un temps de réussite de la tâche très court. Ainsi le poids de ce critère pourra être prédominant par rapport aux autres (*i.e.* critère pondéré avec un poids fort).

4.4 Les évaluateurs

Le choix des « évaluateurs » devra également être réalisé avec soin pour obtenir un panel d'usagers représentatif des catégories d'usagers pour lesquelles les IUI vont être testées. Ainsi, les usagers pourront par exemple être répartis en neuf classes selon les caractéristiques des usagers en RI (*c.f.* tableau 1). Ce classement peut être réalisé par l'intermédiaire d'un questionnaire ouvert par exemple.

		Connaissance du domaine de recherche		
		Néophyte	Intermédiaire	Expert
Connaissance de l'outil informatique (2D/3D...)	Néophyte			
	Intermédiaire			
	Expert			

TAB. 1 – Catégories des usagers des IUI (échelle ouverte).

Les évaluations pourront alors se dérouler avec des jeux de test spécifiques (à la tâche) et identiques pour tous les évaluateurs. À la suite des différentes évaluations, les résultats pourront être synthétisés pour vérifier l'adéquation entre l'IUI, les tâches et les usagers. Ceci peut être réalisé par le calcul de corrélations par exemple.

5 Critères d'évaluation

Dans cette section, nous proposons quelques critères d'évaluation permettant de comparer des IUI respectant les contraintes précédemment citées. La liste de critères proposée n'est pas exhaustive et certains critères peuvent ne pas être adaptés à toutes les interfaces. L'idée étant de faire émerger des relations entre des critères et la réussite ou non d'une recherche d'information. La sélection des critères à prendre en compte doit être effectuée en amont de la phase d'évaluation et doit correspondre à une caractérisation des différentes tâches évaluées. En effet, l'évaluation d'une IUI, pour une tâche donnée, se traduit par la prise en compte d'un certain nombre de critères et éventuellement par une pondération de ces critères.

5.1 Temps nécessaire pour achever la tâche

Ce critère consiste à mesurer le temps mis par l'utilisateur pour achever la tâche qu'il doit réaliser. Ainsi ce temps peut être par exemple le temps nécessaire pour trouver un résultat pertinent pour la requête effectuée. Ce critère est donc mesuré objectivement mais nécessite d'avoir préalablement identifié l'ensemble des résultats pertinents pour la requête. Il peut également s'agir de mesurer, pour certaines tâches, le temps nécessaire à l'utilisateur pour identifier un ensemble (ou la totalité) de documents pertinents (similaires ou non). Dans ce cas, le nombre de résultats à retrouver par l'utilisateur doit être fixé et cohérent par rapport au nombre total de résultats pertinents pouvant être retrouvés. On peut également procéder inversement, en fixant le temps de recherche et en comptant le nombre de résultats pertinents identifiés par l'utilisateur. La mesure de ce critère reste alors toujours objectivement quantifiable.

5.2 Ressources nécessaires pour remplir une tâche

Le temps pour trouver le résultat n'est pas forcément pertinent surtout si l'on considère des usagers peu habitués aux applications informatiques. Ainsi, le nombre d'actions à réaliser ainsi qu'un malus dans le cas d'un retour en arrière pourrait être couplé au temps mis pour achever la tâche.

5.3 Temps pour mémoriser et retourner à un document précédemment identifié

Ce critère mesure le temps nécessaire à l'utilisateur pour qu'il retrouve un document qu'il a précédemment identifié (ou qu'il connaît) dans l'interface proposée. Il peut s'agir d'un résultat que l'on vient de voir dans l'espace des réponses d'une requête (*i.e.* on n'a pas quitté

l'espace des réponses) ou d'un document dont on sait qu'il appartient aux résultats d'une requête (*i.e.* on a quitté l'espace des réponses pour y revenir plus tard).

5.4 Temps de réponse du système

Il s'agit là d'un critère simple mais inévitable. En effet, la rapidité est une notion fondamentale pour l'utilisateur lors d'une tâche de recherche d'information. Le calcul du temps de réponse du système consiste simplement à mesurer le temps écoulé entre la validation de la requête et l'affichage des résultats. On considère ici que l'affichage des résultats correspond à l'affichage du dernier composant graphique de l'interface. Cependant, il n'est plus suffisant de se contenter d'une seule valeur. En effet, les traitements des IUI sont de plus en plus variés. Les deux traitements les plus fréquents sont généralement l'affichage et le calcul d'un *clustering* à la volée. Dans ce cas, il semble important d'être en mesure de différencier le temps de calcul pour ces deux tâches. Cette différence permet d'effectuer une meilleure comparaison des systèmes et d'identifier les raisons de leurs potentielles lacunes.

5.5 Compréhension ou « utilisabilité » de l'interface

Il s'agit d'un critère subjectif pouvant être évalué par des questions posées à l'utilisateur sur le fonctionnement de l'interface (Shneiderman, 1998). Cependant, afin de garantir « l'égalité » des évaluations par rapport à ce critère, les mêmes conditions doivent être respectées : temps d'explication de l'interface, temps de manipulation de l'interface, présence d'une aide en-ligne. Dans le cas des IUI, le succès d'une interface passe bien souvent par une compréhension intuitive de l'interface (*i.e.* sans explications pour les fonctionnalités fondamentales). On peut aussi ne pas chercher à évaluer ce critère, considérant qu'il influe plus ou moins dans l'évaluation d'autres critères tels que le temps de recherche de résultats pertinents.

5.6 Capacités de personnalisation et d'adaptation

Dans un premier temps, ce critère peut-être évalué de manière binaire : l'interface est-elle personnalisable et/ou est-elle adaptative ? La personnalisation concerne la possibilité offerte à l'utilisateur de modifier certains paramètres de l'interface alors que l'aspect adaptatif est relatif à la prise en compte par l'IUI de certaines connaissances qu'elle possède de l'utilisateur. Dans le cas où l'IUI est personnalisable, les points suivants pourraient alors être considérés : quelles parties (ou fonctions) sont personnalisables, quelles compétences sont nécessaires du point de vue de l'utilisateur, quel est le temps nécessaire pour l'utilisateur pour réaliser ces personnalisations, quel est l'apport de la personnalisation... Concernant le critère d'adaptation, on pourrait entre autres vérifier si l'interface exploite les précédentes requêtes formulées par l'utilisateur pour adapter, réorganiser les résultats comme le propose par exemple Lainé-Cruzel (1999). Un autre point consiste à vérifier si l'IUI est en mesure de présenter la représentation de l'utilisateur qu'elle possède. Ce dernier point est prédominant pour une bonne acceptation de la personnalisation proposée. Et là aussi il faudrait pouvoir identifier l'apport de l'adaptation (le résultat est-il plus pertinent ?...). Ce critère de personnalisation et d'adaptation de l'IUI reste cependant trop dépendant de nombreux aspects

(tels que du type d'interface ou de personnalisation proposés) pour tenter une généralisation de l'évaluation de ce critère.

5.7 Réussite de la tâche proposée

L'utilisateur doit juger s'il a réussi ou non à réaliser la tâche qui lui était proposée. Cet achèvement peut être binaire ou être associé à une échelle ouverte (note de 1 à 5 par exemple). L'intérêt de ce critère est qu'il permettra de calculer l'adéquation entre l'IUI, la tâche et le type d'utilisateur par exemple.

5.8 Caractéristiques du résultat de recherche

Les critères à prendre en compte ne sont pas tous uniquement liés à l'utilisateur. En effet, pour une tâche donnée, une IUI peut être efficace pour un nombre limité de documents retrouvés et ne plus l'être pour un nombre plus important. Ainsi, nous pouvons souligner les caractéristiques suivantes :

- nombre de documents retrouvés ;
- longueur des documents (en moyenne et en écart type) ;
- hétérogénéité des thèmes abordés dans les documents ;
- hétérogénéité du contenu des documents (contiennent-ils du texte, des images...) ;
- hétérogénéité de structure (est-ce des documents XML, du texte libre ou des données structurées ?).

Ces différents critères devront être combinés pour mesurer l'adéquation entre l'IUI, la tâche de l'utilisateur et le type d'individu. En effet, les utilisateurs peuvent être caractérisés par deux caractéristiques spécifiques pour la recherche d'information : la connaissance du domaine et la connaissance pratique (des outils informatiques). Une étude souligne l'importance de ces critères qui influent sur la façon dont les utilisateurs recherchent l'information (Hölscher, 2000). Grâce à ces différentes informations nous pouvons donc s'assurer qu'une IUI s'adapte à certains publics et à certaines tâches.

6 Conclusion

Il existe un grand nombre d'outils de recherche d'information notamment disponibles sur le Web. Ces différents outils tentent de répondre au mieux aux attentes des utilisateurs aussi variés que les besoins en information qu'ils tentent d'assouvir. La stratégie actuelle est de faire progresser les outils en termes de fonctionnalités intrinsèques (amélioration du processus d'indexation, adaptation...), mais un aspect important, lié à l'utilisateur, reste la conception et l'intégration de l'IUI. Cette IUI est la passerelle directe entre l'outil de recherche et l'utilisateur au travers de laquelle les documents retrouvés sont présentés. Ces IUI doivent ainsi être bien réfléchies pour permettre à une catégorie d'utilisateurs de remplir des tâches spécifiques. Cependant, à l'heure actuelle nous ne sommes pas en mesure de vérifier si au moins une IUI répond bien aux attentes. Le besoin en évaluation est donc important pour

vérifier ces hypothèses mais également de vérifier si éventuellement une IUI pouvait remplir d'autres tâches non prévues initialement ou si elle était adaptée à d'autres catégories d'utilisateurs. Par ailleurs, cette évaluation permet en outre de réaliser une comparaison et une classification des IUI en fonction des tâches qu'elles permettent de remplir par exemple. Nous proposons dans ce contexte une approche de l'évaluation des IUI au regard de différents critères. Ces critères permettent la caractérisation des tâches que l'on souhaite évaluer. Grâce aux évaluations et aux mesures ainsi calculées au travers des différentes évaluations, une synthèse pour chaque IUI pourra être fournie. Cependant, la phase d'évaluation n'est pas si simple à réaliser et surtout à organiser comme le souligne par exemple Fekete (2004). Nous devons désormais poursuivre notre réflexion sur les indicateurs et la caractérisation des différentes tâches à partir de ceux-ci. Nous devons également approfondir le mode opératoire de cette phase d'évaluation afin d'obtenir et de valider les résultats que nous espérons aussi objectifs que possible.

Références

- Bonnel N., Cotarmanac'h A., et Morin A. (2005a). Gestion et visualisation des résultats d'une requête. *Actes du 3^{ème} atelier Visualisation et Extraction de Connaissances (associé à EGC'05)*, pp. 37–47.
- Bonnel N., Cotarmanac'h A., et Morin A. (2005b). Meaning Metaphor for Visualizing Search Results. *Proceedings of the 9th International Conference on Information Visualisation*, pp. 467–472, IEEE Computer Society.
- Bonnel N., Cotarmanac'h A., et Morin A. (2005c). Visualisation 3D des résultats de recherche : quel avenir ? *Créer, jouer, échanger : expériences de réseaux (Actes de la conférence H2PTM'05 : Hypermedias Hypertexts, Products, Tools and Methods)*, pp. 325–339, Hermes Science Publications.
- Chevalier M. (2002). *Interface adaptative pour l'aide à la recherche d'information sur le web*. Thèse de doctorat, Université Paul Sabatier, Toulouse, France.
- Chevalier M. et Hubert G. (2005). Evaluation d'une interface de restitution de recherche : Quelles conclusions en tirer ? *Actes du 3^{ème} atelier Visualisation et Extraction de Connaissances (associé à EGC'05)*, pp. 15–27.
- Cugini J., Laskowski S., et Sebrechts M. (2000). Design of 3D Visualization of Search Results: Evolution and Evaluation. *Proceedings of IST/SPIE's International Symposium: Electronic Imaging 2000: Visual Data Exploration and Analysis*.
- Fekete J.-D. et Plaisant C. (2004). Les leçons tirées de deux compétitions de visualisation d'information. *ACM IHM 2004*, pp. 7–12.
- Hearst M. et Karadi C. (1997). Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy. *Proceedings of International ACM/SIGIR Conference*, pp. 246–255.
- Hölscher C. et Strube G. (2000). Web search behavior of internet experts and newbies. *Proceedings of the 9th International Conference on the World Wide Web (www9)*.

Lainé-Cruzet S. (1999). ProfilDoc – Filtrer une information exploitable. *Bulletin des bibliothèques de France (BBF)*, 44(5), pp. 60–64.

Robertson G., Mackinlay J., et Card S. (1991). Cone Trees: Animated 3D Visualizations of Hierarchical Information. *Proceedings of ACM CHI'91 Human Factors in Computing Systems Conference*, pp. 189–194, ACM Press New York.

Shneiderman B. (1998). *Designing the User Interface*. Addison-Wesley.

Annexe : récapitulatif

Cette annexe permet de faire un récapitulatif des contraintes et critères proposés. Le tableau 2 constitue en quelque sorte un mémo d'évaluation, récapitulant un ensemble de contraintes à satisfaire afin de pouvoir utiliser certains des critères proposés en vue d'une comparaison avec d'autres systèmes.

CONTRAINTES	CRITÈRES
<ol style="list-style-type: none"> 1. Données 2. Fonctionnalités 3. Jeu de requêtes 	<ol style="list-style-type: none"> 1. Temps nécessaire pour achever la tâche 2. Ressources nécessaires pour accomplir une tâche 3. Temps pour mémoriser et retourner à un document précédemment identifié 4. Temps de réponse du système 5. Compréhension ou utilisabilité de l'interface 6. Capacités de personnalisation et d'adaptation 7. Réussite de la tâche proposée
PRÉALABLES	
<ol style="list-style-type: none"> 1. Définir et caractériser les tâches à évaluer 2. Choisir un panel d'utilisateurs représentatif 	<ol style="list-style-type: none"> 8. Caractéristiques du résultat de recherche

TAB. 2 – Récapitulatif.

Summary

An information retrieval (IR) process has no sense for users without the last step which consists in visualizing the search results. The increase of the importance of search result visualization is responsible of the proposition of many interfaces in the last few years. These interfaces can be based on textual, 2D or 3D metaphors. Although some of these interfaces were evaluated, no effective comparison was carried out in this context due to the lack of constraints on these interfaces and due to the lack of comparison criteria concerning the IR task. So we propose to introduce some evaluation tracks in order to obtain a framework enabling the evaluation and the comparison of various **Information User Interfaces**.

Keywords: Information User Interface, Visualization of Search Results, Information Retrieval System, Evaluation and Comparison Criteria.