HAL open science

Critères d'évaluation pour les interfaces des systèmes de recherche d'information

Nicolas Bonnel, Max Chevalier

▶ To cite this version:

Nicolas Bonnel, Max Chevalier. Critères d'évaluation pour les interfaces des systèmes de recherche d'information. 3ème Conférence en Recherche d'Information et Applications (CORIA'06), LIRIS, Mar 2006, Lyon, France. pp.109-114. inria-00001155

HAL Id: inria-00001155

https://inria.hal.science/inria-00001155v1

Submitted on 22 Mar 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Critères d'évaluation pour les interfaces des systèmes de recherche d'information

Nicolas Bonnel* — **Max Chevalier****,*****

** IRISA / Université de Rennes 1*
Campus universitaire de Baulieu
35042 Rennes Cedex
nicolas.bonnel@irisa.fr

*** IRIT / Université Paul Sabatier*
118 route de Narbonne
31062 Toulouse Cedex
Max.Chevalier@irit.fr

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

Catégorie chercheur

RÉSUMÉ. *La visualisation (ou restitution) des résultats d'une recherche est une étape essentielle dans tout processus de recherche d'information. En effet, les interfaces utilisateur d'information servent de lien entre les utilisateurs et les systèmes de recherche d'information, et permettent donc de donner un « sens » aux résultats pour les utilisateurs. Face à l'importance que prend la visualisation des résultats, de nombreuses interfaces (textuelles, 2D ou 3D) ont été proposées depuis une dizaine d'années. Cependant, bien que des évaluations aient été proposées, elles sont trop rares et peu interprétables (et comparables) du fait de la trop grande hétérogénéité des systèmes et de l'absence d'un cadre strict d'évaluation. Dans cet article, nous proposons donc de définir un ensemble de critères dans le but d'évaluer et de comparer ces interfaces.*

ABSTRACT. *Visualization of search results is an essential step in the Information Retrieval process. Indeed Information User Interfaces are used as a link between users and Information Retrieval Systems, and so they enable to give sense to the results for users. Due to the importance that takes visualization of search results, many interfaces (which can be textual, in 2D or in 3D) have been proposed in the last decade. However, although some evaluations of these interfaces were proposed, they are too rare and not really interpretable (and comparable) due to the big heterogeneity of systems and due to the absence of an evaluation framework. So, in this paper, we propose to define a set of criteria which take users, tasks and system into account.*

MOTS-CLÉS : *interface utilisateur d'information, interface de restitution de résultats de recherche, systèmes de recherche d'information, critères d'évaluation et de comparaison.*

KEYWORDS: *Information User Interface, visualization of search results, Information Retrieval Systems, evaluation and comparison criteria.*

1. Introduction

La manipulation des systèmes de recherche d'information (SRI), et notamment des moteurs de recherche dans le cas de la recherche d'information (RI) sur le Web, se fait généralement au travers d'un frontal permettant à l'utilisateur de spécifier sa requête et d'en visualiser les résultats. C'est pourquoi nous utiliserons désormais, dans cet article, le terme d'**interface utilisateur d'information (IUI)**[1] pour parler de cette étape qui consiste à visualiser les résultats issus d'un SRI.

Dans cet article, nous nous intéressons à l'élaboration de critères relatifs à l'évaluation de ces IUI, qui est une tâche particulièrement délicate et difficile. Le but de ces critères est de caractériser l'impact d'une interface dans le processus de RI. Or, jusqu'ici, l'évaluation était essentiellement orientée « utilisabilité » ; cette évaluation n'est pas suffisante car elle ne permet pas, à notre sens, de mettre en évidence l'impact de l'IUI sur la tâche. C'est pourquoi nous proposons un ensemble de critères nous permettant d'évaluer, de comparer et de classifier les différentes IUI en fonction des tâches, des utilisateurs…

L'intérêt de ces IUI n'étant plus à démontrer [BON 06] nous avons choisi de se focaliser, dans cet article, sur les critères d'évaluation de ces interfaces. Ainsi, la **section 2** fait le point sur l'évaluation actuelle des IUI et permet d'en identifier les limites. Ensuite la **section 3** est consacrée à la proposition de critères d'évaluation. Enfin, la **section 4** donne quelques précisions sur les contraintes à respecter lors d'une telle phase d'évaluation.

2. Évaluation d'une IUI

Actuellement, 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 évaluent et notent l'IUI par le biais d'une série de questions. Ce questionnaire est très souvent inspiré des propositions de [SHN 98] en matière d'évaluation d'interfaces utilisateur, et est très orienté « utilisabilité ». Ce principe a notamment été adopté dans [CHE 05, BON 05]. 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 admettent souvent une réponse subjective, ne sont pas toujours adaptées au contexte des IUI, et ne permettent pas tout simplement d'apprécier l'impact de l'IUI sur le processus de RI [CHE 05]. Dans le même temps, il est aussi important de valider les hypothèses émises lors de la conception de l'IUI (par exemple, l'IUI est-elle réellement destinée à des utilisateurs *lambda* ou permet-elle réellement de remplir telle ou telle tâche ?). Il est donc nécessaire de proposer une évaluation, complémentaire aux évaluations plus orientées « utilisabilité », permettant de caractériser plus finement ces IUI dans leur utilisation en RI.

1. Ce terme a été proposé par Franck Poisson, ex-Directeur Général et fondateur du bureau France de *Google*.

Pour cela, différents critères doivent permettre de mesurer la qualité d'une IUI en fonction d'un contexte spécifique (type d'usager, type de résultat…) de la RI. Nous pouvons alors supposer que ces critères sont principalement liés à l'usager comme le souligne [LAI 99] qui identifie que l'usager peut être caractérisé par trois questions : « Qui est-il ? Que veut-il ? Pour quoi faire ? ». Cependant, pour obtenir une caractérisation fidèle de l'IUI nous pouvons faire émerger trois catégories de critères qui sont liés à l'**usager** (« Qui est-il ? »), à sa **tâche** (« Que veut-il ? Pour quoi faire ? »), et enfin au **système** (notamment aux caractéristiques de l'ensemble des documents obtenus en résultat). Grâce à ces différentes catégories de critères nous pouvons ainsi caractériser au mieux l'impact d'une IUI sur le processus de RI en fonction de différents points de vue.

3. Critères d'évaluation

Dans cette section, nous proposons les critères d'évaluation que nous avons identifiés selon les différents points de vue, permettant de comparer les IUI. La liste de critères proposée n'est pas exhaustive et certains critères peuvent ne pas être adaptés à toutes les interfaces.

3.1. *Critères liés à l'usager*

Il est important de préciser que les critères présentés dans cette partie peuvent être évalués *a priori* lors de la préparation de la phase d'évaluation et notamment lors du choix du panel d'utilisateurs.

Critère 1 : niveau de connaissance du domaine lié à l'information recherchée. Comme le souligne [HOL 00], la réussite d'une recherche d'information dépend entre autres du niveau de connaissance du domaine lié à l'information recherchée. Ce critère est donc très important car il traduit en quelque sorte l'expertise de l'usager dans le thème lié aux informations qu'il recherche.

Critère 2 : niveau de connaissance pratique de l'outil informatique. Ce critère est complémentaire au critère précédent. Il permet de savoir si un usager sait manipuler l'environnement informatique et notamment celui de la RI. Cela peut aller de la connaissance de l'utilisation de la souris et du clavier, du fonctionnement des SRI ou alors des interfaces 3D…

3.2. *Critères liés à la tâche de l'usager*

Critère 3 : réussite de la tâche proposée. L'utilisateur doit juger s'il a réussi ou non à réaliser la tâche qui lui était proposée. L'intérêt de ce critère est qu'il permet de calculer l'adéquation entre l'IUI, la tâche et le type d'utilisateur par exemple.

Critère 4 : temps nécessaire pour effectuer la tâche. Ce critère consiste à mesurer le temps mis par l'utilisateur pour effectuer la tâche qu'il doit réaliser. Contrairement au critère précédent, il ne prend pas en compte le fait que la tâche ait été réussie ou non.

Critère 5 : ressources nécessaires pour effectuer la tâche. Le temps nécessaire pour effectuer une tâche n'est pas forcément gage de qualité surtout si on considère des usagers peu habitués aux applications informatiques. Ainsi, ce critère permet d'introduire un coût de manipulation qui peut être couplé au temps mis pour effectuer la tâche.

3.3. *Critères liés au système*

Dans cette catégorie de critères, nous distinguons deux types de critères : ceux qui sont liés à l'outil de recherche dans sa globalité (système de recherche **et** interface) et ceux qui sont liés à l'ensemble des résultats. Nous commençons par le premier type de critères, c'est-à-dire ceux liés à l'outil de recherche.

Critère 6 : temps de réponse. Il s'agit d'un critère simple et trivial, mais inévitable. En effet, la rapidité est une notion fondamentale pour l'utilisateur lors d'une tâche de recherche d'information. Ce temps correspond, dans la phase d'évaluation, au temps entre la soumission de la requête et l'affichage effectif des résultats.

Critère 7 : compréhension ou « utilisabilité » de l'interface. Il s'agit d'un critère subjectif. Il repose notamment sur des questions posées à l'utilisateur sur le fonctionnement de l'interface [SHN 98]. Cependant, afin de garantir « l'égalité » des évaluations par rapport à ce critère, les mêmes conditions de manipulation de l'IUI doivent être respectées.

Critère 8 : temps pour retrouver 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. Ce critère prend donc en compte la capacité de mémorisation de l'utilisateur, qui peut notamment être améliorée dans le cas de métaphores de visualisation spatiales (utilisation de notre mémoire spatiale).

Critère 9 : capacités de personnalisation et d'adaptation. Ce critère indique si l'interface est personnalisable et/ou adaptative.

Les critères qui suivent sont liés à l'ensemble des documents retrouvés.

Critère 10 : nombre de documents retrouvés. Ce critère indique le nombre de documents retournés par le SRI.

Critère 11 : longueur des documents retrouvés. Ce critère concerne la longueur des documents retrouvés. Par exemple, dans le cas de textes, cette longueur peut correspondre au nombre de mots présents dans le document.

Critère 12 : homogénéité des documents. Ce critère indique si les documents retrouvés forment un ensemble homogène (ou au contraire un ensemble hétérogène), du point de vue du thème par exemple. Ainsi, on peut mettre en évidence si l'IUI permet à l'usager de s'orienter rapidement vers les documents liés au thème de la recherche.

Critère 13 : type de documents. Ce critère indique quels types d'informations sont restituées par le SRI.

Critère 14 : performances intrinsèques du système. Ce critère correspond aux performances intrinsèques du système, dont dépend notamment la qualité des résultats.

Le choix des critères à prendre en compte pour l'évaluation dépend de la tâche à réaliser ainsi que du contexte. Ainsi, chacune des tâches identifiées dans la phase d'évaluation doit être associée à ces critères organisés selon leur importance dans la tâche. Par exemple, si l'usager recherche un document précis, le critère de temps de réponse peut être prédominant sur le reste. Ce critère est encore plus important pour la tâche si on la considère dans un contexte tel que la RI sur le Web. Par ailleurs, certains critères semblent corrélés (par exemple le niveau d'expertise et le temps nécessaire pour effectuer la tâche). Toutefois, nous les considérons comme indépendants pour éviter tout problème d'interprétation et surtout de comparaison des résultats de différentes évaluations. Les différents contextes de recherche peuvent aussi être caractérisés par certains de ces critères (éventuellement pondérés) afin de mesurer l'adéquation d'une IUI à un contexte spécifique.

4. La phase d'évaluation

Afin de pouvoir évaluer et comparer efficacement différentes IUI (en utilisant notamment les critères définis à la section précédente), certaines contraintes doivent être respectées lors de la phase d'évaluation. Ces contraintes doivent même être prises en compte dès la conception de l'IUI afin de garantir leur respect. Ainsi l'IUI doit être :

– **indépendante** du système de recherche (*i.e.* du calcul des résultats). Pour cela, il est nécessaire que l'IUI accède aux données *via* des interfaces (ou API[2]) riches en termes de fonctionnalités et libres d'utilisation.

– **modulable**, 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 jeux de requêtes doivent permettre de couvrir les différentes tâches pour lesquelles on souhaite évaluer l'IUI ; ce qui implique d'avoir préalablement défini et caractérisé ces tâches.

Le choix des « évaluateurs » doit également être réalisé avec soin pour obtenir un panel d'utilisateurs représentatif des catégories d'usagers pour lesquelles on souhaite évaluer l'IUI, mais aussi représentatif pour les tâches définies.

2. *Application Programming Interface*

5. Conclusion

Il existe un grand nombre d'outils de recherche d'information, qu'il s'agisse de SRI spécialisés ou de moteurs de recherche sur le Web. Ces différents outils tentent de répondre au mieux aux attentes des utilisateurs qui sont 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 (telles que l'amélioration du processus d'indexation ou l'adaptation du système de recherche). Toutefois, une fonctionnalité liée à l'usager reste la conception de l'interface. Cette IUI est la passerelle directe entre l'outil de recherche et l'usager, au travers de laquelle les documents retrouvés sont présentés. Ces IUI doivent ainsi être bien réfléchies pour permettre à une catégorie d'usagers 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 pour vérifier si éventuellement une IUI peut remplir d'autres tâches non prévues initialement ou si elle est adaptée à d'autres catégories d'usagers. 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. La phase suivante est de modéliser au mieux les contextes en fonction de ces critères et de réaliser une phase d'évaluation réelle.

6. Bibliographie

[BON 05] BONNEL N., COTARMANAC'H A., MORIN A., « Visualisation 3D des résultats de recherche : quel avenir ? », *Créer, jouer, échanger : expériences de réseaux (Actes de la conférence H^{2}PTM'05 : Hypermedias Hypertexts, Products, Tools and Methods)*, Hermes Science Publications, 2005, p. 325-339.

[BON 06] BONNEL N., CHEVALIER M., « Évaluation des Interfaces Utilisateur d'Information », *Actes du 4e atelier Visualisation et Extraction de Connaissances - EGC'06*, 2006, p. 23-38.

[CHE 05] CHEVALIER M., HUBERT G., « Évaluation d'une interface de restitution de recherche : Quelles conclusions en tirer ? », *Actes du 3e atelier Visualisation et Extraction de Connaissances - EGC'05*, 2005, p. 15-27.

[HOL 00] HOLSCHER C., STRUBE G., « Web Search Behavior of Internet Experts and Newbies », *Proceedings of the 9th International Conference on the World Wide Web (www9) - Computer Networks*, vol. 33, n° 1-6, 2000, p. 337-346, North-Holland Publishing Co.

[LAI 99] LAINÉ-CRUZEL S., « ProfilDoc - Filtrer une information exploitable », *Bulletin des bibliothèques de France (BBF)*, vol. 44, n° 5, 1999, p. 60-64.

[SHN 98] SHNEIDERMAN B., *Designing the User Interface*, Addison-Wesley, 3e édition, 1998.