La pertinence des outils d'experts au service des non-experts en recherche d'informations

Jérôme Dinet

▶ **To cite this version:**

Jérôme Dinet. La pertinence des outils d'experts au service des non-experts en recherche d'informations. Revue de l'EPI (Enseignement Public et Informatique), 2000, 99, pp.57-68. edutice-00001183

HAL Id: edutice-00001183
https://edutice.hal.science/edutice-00001183

Submitted on 16 Nov 2005

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.

LA PERTINENCE DES OUTILS D'EXPERTS AU SERVICE DES NON-EXPERTS EN RECHERCHE D'INFORMATIONS : UN EXEMPLE AVEC LES OPÉRATEURS BOOLÉENS

Jérôme DINET

INTRODUCTION

Parmi les activités documentaires fortement préconisées et régulièrement pratiquées dans les établissements scolaires, la recherche d'informations primaires et secondaires (dans ce cas, ce sont des références qui sont accessibles) est l'une des plus importantes. Face à l'accroissement de la masse documentaire au sein même des établissements scolaires, à la diversification des supports et à la possibilité d'interroger des fonds distants physiquement, le recours à des outils informatiques de recherche s'est vite révélé nécessaire. Ces outils de recherche (moteurs de recherche, annuaires thématiques et logiciels spécifiques) sont généralement créés par des spécialistes de la documentation, des informaticiens, des graphistes…, bref des experts dans leur domaine. Or, le paradoxe réside dans le fait que ce sont essentiellement des non-experts de ces mêmes disciplines, tels les élèves des niveaux primaires et secondaires, qui utilisent ces outils.

1. LES DIFFÉRENCES ENTRE EXPERTS ET NON-EXPERTS EN RECHERCHE D'INFORMATIONS

Qu'est-ce qui distingue une personne considérée comme experte d'une autre personne considérée comme non-experte, tel un élève, dans le cas de recherche d'informations dans des environnements informatiques ?

Selon Suchman (1987), dans une perspective cognitiviste générale, une personne est dite experte en recherche d'informations lorsqu'elle est

capable d'utiliser ses connaissances et son expérience pour distinguer des informations pertinentes des informations non pertinentes. Le problème est que la distinction entre ces deux valeurs d'une information (pertinente versus non-pertinente) est très difficile à définir (pour une synthèse sur la notion de pertinence des informations, voir Mizzaro, 1998). De plus, l'expertise concerne d'autres étapes et dimensions que la simple évaluation de la pertinence des informations.

En reprenant les travaux de Marchionini (1995), les différences entre un expert et un non-expert en recherche d'informations dans des environnements électroniques relèvent de quatre dimensions :

- les capacités cognitives générales : de manière générale, un non-expert, tel un jeune élève, possède des capacités mnésiques et intellectuelles inférieures à celles d'un expert adulte ;
- l'expertise du domaine thématique de la recherche : plus un individu possède de connaissances sur le thème de sa recherche, plus il évalue vite et efficacement la pertinence des informations trouvées et plus ses requêtes documentaires sont pertinentes. Par exemple, un étudiant en médecine évalue plus vite qu'un étudiant en lettres la pertinence d'une information trouvée dans une base de données médicales. De même, l'étudiant en médecine utilisera des requêtes documentaires plus pertinentes, plus spécifiques et plus précises pour mener sa recherche ;
- l'expertise du système : certaines expériences (par exemple, Kuhlthau, 1999) montrent qu'avec l'expérience et la pratique, un individu modifie ses stratégies de recherche d'informations puisqu'il connaît de mieux en mieux la structure du corpus informationnel, le contenu informationnel et les outils dont il dispose. Par exemple, bien connaître l'organisation des informations dans un cédérom réduit le temps de recherche dans ce même cédérom ;
- l'expertise en recherche d'informations : ce type d'expertise est le plus long à acquérir, puisqu'il concerne les méthodes de recherche d'informations indépendamment des environnements dans lesquels la recherche est effectuée.

Sur la base des dimensions brièvement explicitées ci-dessus, pour Marchionini, Dwiggins, Katz & Lin (1993), toutes les étapes d'une recherche d'informations sont susceptibles d'être réalisées de manière différente selon le niveau d'expertise de l'individu :

- l'identification du besoin d'informations et la compréhension exacte de ce qui est attendu ;
- la définition du problème à résoudre et la planification de la recherche (« quels mots vais-je utiliser pour mener ma recherche ? » ; « que dois-je trouver ? » ; etc.) ;
- le choix du système dans lequel la recherche doit être menée (par exemple, « dois-je utiliser une encyclopédie papier ou une encyclopédie électronique ? ») ;
- la formulation des requêtes documentaires (« quels mots dois-je écrire pour obtenir des résultats pertinents ? » ; « sur quelle catégorie proposée par l'annuaire dois-je cliquer pour trouver des sous-catégories pertinentes ? » ; etc.) ;
- la supervision de son comportement de recherche (ceci suppose un contrôle de sa propre activité, ce qui correspond globalement à un comportement « méta-cognitif ») ;
- le traitement des résultats de la recherche fournis par l'outil et l'évaluation de leurs pertinences respectives.

Bref, il existe de multiples différences entre un expert et un non-expert en recherche d'informations. Mais, après tout, pourquoi les outils créés par les uns se révéleraient inutiles pour les autres ?

2. LES OPÉRATEURS BOOLÉENS : UNE AIDE PERTINENTE ?

Dans le cadre de la recherche d'informations, parmi les outils créés par des experts et censés aider les non-experts, nous pouvons citer les opérateurs booléens. Quel que soit l'outil utilisé, rechercher des informations ou des documents nécessite d'interroger une base de données. Cette interrogation se réalise grâce à la production (à l'aide du clavier alphanumérique le plus souvent) de termes de recherche, aussi appelés requêtes. Aujourd'hui, la plupart des outils de recherche d'informations ou documentaire permet l'emploi d'opérateurs booléens pour effectuer la recherche. Ainsi, les moteurs de recherche et annuaires thématiques disponibles sur le World Wide Web proposent cette fonctionnalité, le plus souvent en mode « recherche avancée » ou encore « recherche experte ». Ces opérateurs, peuvent-ils être utiles à des personnes non-expertes en langages documentaires ?

2.1. Les contextes d'apparition et d'utilisation de la logique booléenne

Rappelons brièvement le contexte dans lequel sont apparus les opérateurs booléens. Les opérateurs booléens reposent sur la logique créée par Georges Boole au XIXe siècle (1815-1864, mathématicien anglais, créateur de l'algèbre logique et du calcul binaire). La logique booléenne a été introduite dans les interfaces d'interrogation des systèmes d'informations dans les années 1950, c'est-à-dire à un moment où :

- très peu de personnes utilisaient l'ordinateur ;
- les capacités de stockage et de traitement de l'information étaient très limitées ;
- les ordinateurs étaient relativement chers et les systèmes de traitement de l'information encore plus chers ;
- les systèmes d'informations et documentaires étaient réservés à des spécialistes de la documentation.

Or, le contexte actuel est bien différent. Le tableau 1 résume les comparaisons entre le contexte d'apparition des opérateurs booléens dans les systèmes informatiques et le contexte actuel, sur les mêmes dimensions.

Tableau 1 :
Comparaison du contexte d'apparition des opérateurs booléens dans les systèmes d'informations et du contexte actuel

Dimensions	***Avant 1960***	En 2000
Nombre d'utilisateurs	Peu nombreux	Très nombreux
Capacités des systèmes	Faibles et très limitées	Importantes et croissantes
Coût des systèmes	Onéreux	Abordables
Disponibilité des systèmes	Peu accessibles	Accessibles
Expertise des utilisateurs	Élevée	Très hétérogène

Aujourd'hui, de très nombreuses personnes utilisent les outils informatiques qui sont devenus beaucoup plus accessibles, tant du point de vue financier que de celui de leur présence dans nos environnements

(familiaux, scolaires, professionnels). Et la plupart de ces utilisateurs ne sont pas des spécialistes de la documentation.

2.2. Le langage « booléen » et le langage « naturel »

Depuis la fin des années 1970, de nombreux travaux ont tenté de créer des langages d'interrogation utilisables facilement par des non professionnels (par exemple, Borgman, 1986 ; Cooper, 1988 ; Green, Devlin, Cannata & Gomez, 1990 ; Young & Schneiderman, 1993). Les conclusions sont multiples et parfois difficiles à interpréter. Néanmoins, les auteurs semblent d'accord sur un point : la présence d'opérateurs booléens sur les interfaces d'interrogation n'est d'aucune utilité pour les non-professionnels de la documentation et les perturbe souvent. Pourquoi ?

Parmi les hypothèses explicatives, il est raisonnable de retenir la suivante : les opérateurs booléens sont très peu et incorrectement utilisés parce que la logique impliquée par ces opérateurs ne correspond pas à la logique utilisée habituellement par un individu. En effet, le raisonnement induit par les connecteurs logiques est différent selon que ces connecteurs logiques sont utilisés en tant qu'opérateurs booléens ou en tant qu'opérateurs dans le langage quotidien : en d'autres termes, tout dépend du contexte dans lequel sont utilisés ces connecteurs.

A titre d'exemple, l'opérateur ET (ou ses correspondants étrangers tels que AND) implique une inclusion dans le langage naturel alors qu'il implique une exclusion dans le langage documentaire. Ainsi, au restaurant, lorsqu'une personne demande du chocolat ET un café, elle espère généralement les deux. Par contre, en « langage documentaire », demander « chocolat ET café » correspond à une restriction du champ de réponses : dans ce cas, seul le croisement des deux ensembles « chocolat » et « café » est attendu. Dans le cas d'une recherche documentaire informatisée, seules les références de documents contenant des informations sur les deux concepts simultanément sont attendues.

De même, l'opérateur OU implique une inclusion dans le langage documentaire alors qu'il implique une exclusion (donc une restriction) dans le langage naturel. Parfois, dans la vie courante, il faut choisir : boire OU conduire. Dans le « langage documentaire », mettre un OU entre deux termes signifie que l'on effectue une recherche simultanée sur les deux concepts, c'est-à-dire que « l'on souhaite les deux en même temps ». Le tableau 2 résume les correspondances pour les deux opérateurs (ET et OU) selon le contexte d'utilisation.

Tableau 2 :
Correspondances entre le langage documentaire et le langage « naturel » pour les connecteurs logiques ET et OU

	Langage documentaire	***Langage « naturel »***
ET	Exclusion	Inclusion
OU	Inclusion	Exclusion

Alors, que dire lorsqu'un élève doit mener une interrogation documentaire en combinant des ET et des OU ? Sans parler des syntaxes d'interrogation spécifiques à certains outils. Par exemple, avec quelques annuaires thématiques, la requête « France + île » ne signifie aucunement que la recherche porte sur l'association des deux concepts. En fait, cette requête signifie que le concept « île », précédé du signe additionnel « + », doit absolument apparaître dans toutes les réponses fournies par l'outil.

En considérant que les opérateurs booléens sont des objets finalisés, c'est-à-dire créés en vue de fonctions et finalités spécifiques et si nous reprenons certains éléments de la définition établie par Tricot & Tricot (2000), à savoir que :

- l'utilisabilité d'un objet finalisé désigne la possibilité d'utilisation de cet objet ;
- l'utilité d'un objet finalisé désigne la possibilité d'atteindre un but visé avec cet objet.

alors, nous pouvons nous poser les deux questions suivantes :

- les opérateurs booléens sont-ils effectivement utilisés par des utilisateurs non-experts ?
- les opérateurs jouent-ils leur « rôle » à savoir aider les utilisateurs non-experts à formuler leurs requêtes documentaires ?

Afin de répondre à ces deux questions, une expérience a été réalisée. Dans cette expérience, je me suis intéressé à l'utilisation effective d'opérateurs booléens par des utilisateurs de différents niveaux d'expertise quant à la recherche d'informations.

3. UNE APPROCHE EXPÉRIMENTALE

Dans une étude expérimentale, des participants avaient pour consigne de produire des requêtes en vue de rechercher des références bibliographiques dans une base de données informatisée.

Le thème de la recherche. Le thème de la recherche était « les armures des chevaliers ». Ce thème, directement relié à l'histoire médiévale, avait été choisi de telle sorte que tous les participants puissent comprendre et se « représenter » ce dont il s'agissait.

Les participants. Sur la base des réflexions de Marchionini (1995), trois niveaux d'expertise en recherche documentaire ont été distingués pour constituer les groupes de participants. Dix-huit participants étaient des étudiants préparant le CAPES de documentation, vingt-deux participants étaient des documentalistes en poste depuis au moins dix ans et soixante-sept participants étaient des étudiants ne préparant pas le CAPES de documentation. Les étudiants ne préparant pas le CAPES sont dénommés « novices », les étudiants préparant le CAPES de documentation sont considérés comme « intermédiaires » et les documentalistes en poste sont considérés comme « experts ». Parmi les participants « novices », trente-huit étaient en lettres modernes et vingt-neuf étaient en histoire. Les seconds devaient donc a priori posséder plus de connaissances sur le thème de recherche (« les armures des chevaliers »). Qu'ils soient en lettres modernes ou en histoire, tous appartenaient à la même université. De plus, tous étaient francophones et avaient une moyenne d'âge sensiblement équivalente (24,5 ans pour les étudiants en histoire et 25,9 ans pour les étudiants en lettres modernes) et tous étaient inscrits dans le même niveau du cursus universitaire français (année de maîtrise). Enfin, tous les participants « novices » ont déclaré avoir une faible connaissance des techniques de recherche documentaire et du maniement d'ordinateurs en général. Le tableau 3 résume les principales caractéristiques des participants.

Tableau 3 :
Principales caractéristiques des participants à l'expérience

	Groupe « Novices »		Groupe « Intermédiaires »	Groupe « Experts »
	Etudiants en histoire	Etudiants en lettres modernes	Etudiants préparant le CAPES de documentation	Documentalistes en poste
Nombres	29	38	18	22
Âges moyens	24,5 ans	25,9 ans	23,9 ans	35,4 ans
Langue maternelle	***Français***			

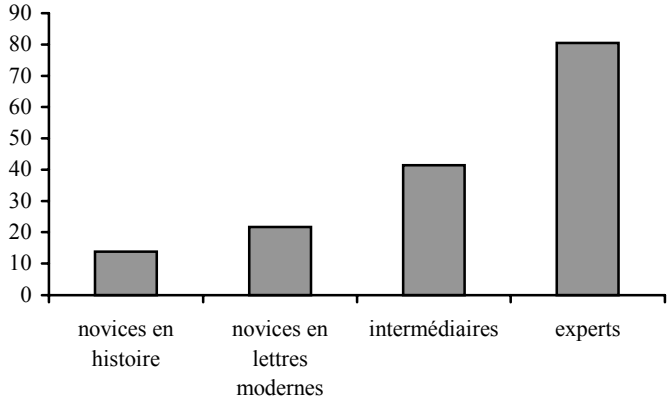
Le matériel et la procédure. La passation de l'étude était individuelle. Le thème (« les armures des chevaliers ») était présenté au participant qui devait aussitôt écrire le(s) terme(s) avec le(s)quel(s) il allait interroger la base de données informatisée de manière à obtenir des références bibliographiques relatives au thème. La production des requêtes devait être réalisée sur une copie d'écran reprenant exactement un écran de recherche prototypique d'écrans de recherche avancée existants (Figure 1). Aussi, les participants avaient à leur disposition les opérateurs booléens ET, OU et SAUF et pouvaient s'aider des parenthèses. Le temps pour réaliser cette activité était libre.

Figure 1 : *Copie d'écran utilisée par les participants pour produire leurs requêtes*

(CHERCHER	……	OU	……)
ET	……	OU	……)
ET	……	OU	……))
(SAUF	……	OU	……)

Les résultats. Puisque c'est l'utilisation des opérateurs booléens qui représente l'objet d'étude, regardons si ces derniers sont présents dans les requêtes documentaires produites par les participants. La figure 2 présente les fréquences des requêtes (en pourcentages) qui sont produites par les participants et qui comportent au moins un opérateur booléen.

Figure 2 : Fréquences (exprimées en pourcentages) des requêtes comportant au moins un opérateur booléen



Les résultats montrent que la fréquence d'utilisation des opérateurs booléens augmente avec l'expertise en recherche documentaire des participants : les « novices » utilisent peu les opérateurs booléens ; les « intermédiaires », qui sont des étudiants préparant le CAPES de documentation, les utilisent plus fréquemment ; les « experts », qui sont des documentalistes en poste depuis au moins dix ans, les utilisent dans plus de 80% de leurs requêtes.

4. DISCUSSIONS ET IMPLICATIONS PÉDAGOGIQUES

Les résultats obtenus semblent donc confirmer les résultats de travaux antérieurs : les opérateurs booléens ne sont effectivement pas utilisés par des utilisateurs non-experts en recherche d'informations.

De plus, l'hypothèse selon laquelle les participants « novices » utilisent peu les opérateurs booléens parce qu'ils n'y pensent pas semble invalidée puisque tous les participants de notre expérience disposaient de la présence des opérateurs booléens et des parenthèses.

En d'autres termes, notre expérience tend à démontrer que certains outils créés par des experts en vue d'aider des non-experts sont peu utilisés. Selon moi, ces outils sont peu utilisés parce que peu

utilisables sans une formation explicite à ces outils. En effet, les participants de l'expérience qualifiés « d'intermédiaires » sont formés aux techniques d'utilisation des opérateurs booléens. C'est cet apprentissage qui explique certainement la fréquence plus grande d'utilisation de ces outils. Cette fréquence d'utilisation croît d'ailleurs avec « l'habitude » à manipuler ces outils puisque les documentalistes en poste depuis au moins dix ans les utilisent très majoritairement.

Alors, faut-il enseigner l'utilisation des opérateurs booléens ? Aujourd'hui, les avis sont partagés et deux points de vue diamétralement opposés semblent exister :

1- Apprentissage des outils. Il est nécessaire d'initier les élèves aux techniques et outils de recherche d'informations, dont les opérateurs booléens. Certains enseignants vont jusqu'à proposer des initiations aux langages documentaires (tels que Motbis) ou encore aux thesauri. Ainsi, l'utilisateur pourrait maîtriser les outils mis à sa disposition. La logique booléenne s'apprend. Certes. Mais, quelle est l'utilité d'apprendre une logique (c'est-à-dire un mode de raisonnement) qui n'est applicable que dans certaines situations, à savoir des interrogations documentaires ? Et surtout, comment expliquer à des adultes non professionnels de la documentation et de surcroît à des élèves, qu'il existe plusieurs types de logiques selon les contextes d'utilisation ?

2- Apprentissage des méthodes. Les élèves doivent prioritairement être initiés aux méthodes inhérentes à la recherche d'informations, indépendamment des outils. L'idée des tenants de cette position est que ce sont surtout les procédures et attitudes mentales qui sont importantes à acquérir puisqu'elles pourraient être transposées à tous les types d'outils. Ainsi, l'utilisateur ne serait pas « assujetti » à un outil. Cela permettrait de faire prendre conscience à l'élève que c'est lui qui dirige sa recherche ainsi que les outils et non l'inverse. Mais, est-il pertinent de faire travailler l'élève dans l'abstraction d'outils lors d'enseignements ne portant que sur la recherche d'informations informatisée ? En effet, même s'il existe des méthodes applicables à tous les outils, ceux-ci possèdent des structures, systèmes et interfaces qui impliquent une nécessaire contextualisation des méthodes acquises.

Comme j'ai essayé de le démontrer, chacun de ces deux points de vue est basé sur des objectifs différents qui présentent certains avantages mais également certains inconvénients. De manière générale, même si l'apprentissage de méthodes est à privilégier, cet apprentissage ne peut pas s'effectuer indépendamment d'outils. Si les deux points de vue sont

considérés comme étant deux extrêmes d'un même continuum, alors il me semble que chaque séquence pédagogique liée à la recherche d'informations ou documentaire devrait être positionnée le long de ce continuum, selon des objectifs pédagogiques exacts et préalablement définis.

Jérôme DINET
Laboratoire LaCo, Université de Poitiers
CRDP de Poitou-Charentes
jerome.dinet@mshs.univ-poitiers.fr

QUELQUES RÉFÉRENCES

BORGMAN, C.L. (1986). Why are online catalogs hard to use ? Lessons learned from information retrieval studies. *Journal of the American Society for Information Science*, 37, 387-400.

COOPER, W.S. (1988). Getting beyond Boole. *Information Processing and Management*, 24(3), 241-248.

GREEN, S.L., DEVLIN, S.J., CANNATA, P.E. & GOMEZ, L.M. (1990). No IFs, ANDs or Ors : A study of database querying. *International Journal of Man-Machine Studies*, 32, 303-326.

KUHLTHAU, C.C. (1999). The role of experience in the information search process of an early career information worker : Perceptions of uncertainty, complexity, construction, and sources. *Journal of the American Society for Information Science*, 50(5), 399-412.

MARCHIONINI, G., DWIGGINS, S., KATZ, A. & LIN, X. (1993). Information seeking in full-text end-user-oriented search systems : The role of domain and search expertise. *Library and Information Science Research*, 15, 35-69.

MARCHIONINI, G. (1995). Information Seeking in Electronic Environments. Cambridge : Cambridge University Press.

MIZZARO, S. (1998). Relevance : The whole history. In T.B. Hahn & M. Buckland (Eds.), Historical Studies in *Information Science* (pp.221-244). Medford, N.J. : Information Today, Inc.

SUCHMAN, L.A. (1987). *Plans and situated actions : The problem of human machine communication*. New-York : Cambridge University Press.

TRICOT, A. & TRICOT, M. (2000). Un cadre formel pour interpréter les liens entre utilisabilité et utilité des systèmes d'information (et généralisation à l'évaluation d'objets finalisés). *Colloque Ergo-IHM 2000*, Biarritz, 3-6 octobre.

YOUNG, D. & SCHNEIDERMAN, B. (1993). A graphical Filter / Flow representation of boolean queries : A prototype implementation and evaluation. *Journal of the American Society for Information Science*, 44(6), 327-339.