

---

## Apport des services Web dans l'amélioration de l'accès à l'information sur le Web ?

Une étude comportementale des internautes.

**Christian Belbèze\* — Chantal Soulé-Dupuy\*\***

*\* Systemat*

13 rue Andre Villet  
31405 Toulouse Cedex 4  
christian@belbeze.com

*\*\* Institut de Recherche en Informatique de Toulouse - Université Toulouse 1*  
2 Rue du Doyen-Gabriel-Marty,  
F-31042 Toulouse Cedex 9  
chantal.soule-dupuy@univ-tlse1.fr

---

**RÉSUMÉ.** La difficulté à trouver de l'information sur le Web croît et ceci même pour les plus experts des utilisateurs. Afin de mieux comprendre comment recherchent les internautes, nous avons observé cinq adultes et quatre enfants. Des protocoles d'observation différents, présentant tous deux des recherches imposées et libres, ont été définis pour les enfants et les adultes. Nous avons pu définir un certain nombre de comportements, d'attitudes et de difficultés. Le résultat de ces observations ainsi que l'analyse des comportements sont présentés dans ce papier dans le but d'introduire la contribution des services Web dans l'aide à l'accès à l'information et à la recherche sur le Web.

**ABSTRACT.** The difficulty of finding information on the Web grows and this even for the most experts. In order to better understand how search the Net surfers, we observed five adults and four children. Two different protocols of observation, both presenting imposed and free searches, were defined for the children and the adults. Thus, we could define a certain number of behaviours, attitudes and difficulties. The result of these observations as well as the analysis of the behaviours are presented in this paper in order to introduce the contribution of Web services in the improvement of information access and Web searching.

**MOTS-CLÉS :** Accès à l'information, Recherche sur le Web, Comportement des internautes, Etude comportementale, Services Web.

**KEYWORDS** Information access, Web searching, Internet surfer behaviour, behavioural study, Web Services.

---

## **1. Introduction**

Alors même que le World Wide Web (Web) offre de plus en plus de contenu, trouver une réponse ou une information devient de plus en plus difficile. Avant de penser à modifier les solutions de recherche proposées aux internautes, ils nous a semblent important de les observer. Les moteurs de recherche sont classiquement des outils ne guidant pas l'internaute. En fonction de sa connaissance des mécanismes de ces moteurs et de ses propres compétences (à la fois sur le sujet recherché et sur le format de l'information), l'internaute fournit un référentiel de mots clés qui lui est propre, le plus souvent différent de celui (ou de ceux) utilisé(s) par les auteurs des informations. Dans une requête sans résultat, par exemple, un internaute aguerri et parlant anglais va rapidement élargir sa recherche par l'usage de mots clés dans cette langue. Mais ce cas extrême est rare dans un contexte grand public.

Les moteurs de recherche acceptent n'importe quel référentiel de mots clés et retournent le plus souvent des milliers ou des millions de sites et c'est là bien sûr que commencent les difficultés pour les internautes. Quelles adéquations et quelle pertinence représentent ces réponses ? Comment nos internautes décodent-ils les quelques lignes proposées pour déterminer le contenu de la page ? Comment lisent-ils ensuite ces pages une fois qu'ils les ont sélectionnées. Pour répondre à ses questions et nous aider à imaginer le Web de demain et ses systèmes de recherche de l'information, nous avons sélectionné un groupe de neuf personnes ayant de huit à soixante douze ans dont cinq de sexe féminin, dans des milieux socio-culturels différents et nous les avons observés.

Au travers des observations effectuées et des résultats obtenus grâce aux protocoles mis en place, notre but est de proposer des solutions fondées sur le comportement des internautes et destinées à améliorer l'accès à l'information. Ainsi, notre travail vise à considérer la contribution des services Web pour aider les internautes dans la recherche d'information.

Dans ce papier, nous relatons les principaux travaux liés à l'étude du comportement des utilisateurs en situation de recherche d'information sur le Web. Ensuite nous présentons nos expérimentations, ainsi que les résultats des observations effectuées. Dans la dernière partie, nous décrivons l'architecture à base de services Web que nous avons proposée afin d'aider les internautes à rechercher de l'information, à lire et à analyser les résultats et enfin à naviguer.

## **2. Contexte et état de l'art de la recherche d'information sur le Web**

Depuis les débuts des systèmes documentaires, et en particulier depuis les travaux entrepris par Gerald Salton (Salton, 1970) (Salton et McGill, 1983), la recherche documentaire a énormément évolué. Les évolutions n'ont pas tant porté sur les modèles de recherche d'information eux-mêmes, mais plutôt sur les contextes d'utilisation des systèmes de recherche d'information, le Web ayant bouleversé les

pratiques et les besoins en ce qui concerne l'accès à l'information. Les principales caractéristiques en sont une diversification des utilisateurs comme des sources d'information. Les différentes formes d'hétérogénéité qui en découlent accroissent la nécessité de personnaliser, soit l'accès à l'information (afin de rendre les systèmes de recherche d'information accessibles à tous), soit l'information restituée elle-même (afin de la rendre compréhensible et interprétable par l'utilisateur), soit les deux.

De nombreuses études ont mené à la même conclusion, qui est que les systèmes de recherche documentaire (plus particulièrement sur le Web) fournissent des résultats trop décevants (Jansen et Al, 2000) (Savoie et Picard, 2001), (NachMias et Gilad, 2002), (Laurie, 2005). Les modèles traditionnels de recherche documentaire s'avèrent inefficaces en raison de la masse croissante d'information. Même les compagnes d'évaluation, comme TREC (<http://trec.nist.gov/>), CLEF ([www.clef-campaign.org](http://www.clef-campaign.org)), montrent les limites de la faisabilité des expérimentations à grande échelle. Les améliorations pourraient alors s'appuyer sur de nouvelles architectures intégrant de nouvelles technologies, telles que des Services Web, et sur les recherches menées dans le contexte du Web sémantique (<http://www.w3.org>) ou du nouveau Web (cf. le Web 2.0 à <http://www.web2fr.com/>).

### 3. Les expérimentations

Cette section décrit une série d'expérimentations réalisées grâce à la participation active de neuf utilisateurs ayant accepté de prendre part à des protocoles d'observation de pratiques de recherche d'information sur le Web. Nous avons synthétisé ici les résultats des observations des comportements des internautes selon les différents protocoles. Tous les détails (mesures, observations et vidéos réalisées) peuvent être consultés à l'adresse suivante : <http://sissiprojet.free.fr>.

#### 3.1. Les protocoles d'observation

Deux protocoles ont été définis, un pour les cinq adultes et un pour les quatre enfants. Les deux protocoles mélangent des sujets de recherche imposée et des sujets de recherche libre. Une recherche est un ensemble de tâches de lecture et un ensemble de requêtes utilisant des moteurs de recherche. Nous considérerons dans ces expérimentations que chaque « clic » d'envoi d'une requête via un moteur de recherche correspond à une nouvelle recherche.

##### 3.1.1. Les sujets de recherche

**Protocole adulte.** Le protocole adulte se compose de cinq sujets de recherche, trois imposés (une valeur : « *La distance entre deux villes françaises* », un texte : *les paroles de la chanson « All Blues* », et un document : *la partition de l'hymne roumain*) et deux libres :

**Protocole enfant.** Le protocole enfant comprend deux sujets de recherche, un imposé (*La durée de vie d'un dauphin*) et un libre.

Une *recherche libre* consiste à fournir un sujet de recherche sur lequel, autant que possible, l'enfant ou l'adulte devra trouver une réponse à une question ou un document dans un format particulier.

### 3.1.2. Paramètres et indicateurs à observer

A la fin de chaque recherche, soit pour chacune des observations (résolues ou pas), chaque sujet (utilisateur) doit attribuer une note (entre 0 et 10, 0 signifiant « très mal » et 10 « excellent ») à chacun des quatre paramètres suivants :

- *Capacité à comprendre l'information visitée* (ressenti sur la clarté des documents visités) ;

- Intérêt pour les sites rencontrés (ressenti sur la qualité informative sites) ;

- Longueur subjective de la recherche (ressenti du temps passé sur Internet) ;

- Sentiment général (ressenti sur le plaisir éprouvé à surfer sur le Web).

Pour chaque recherche, nous mesurons un certain nombre d'indicateurs définis dans le contexte spécifique du Web pour étudier le comportement des utilisateurs. Ces indicateurs sont les suivants :

- *Temps complet de l'observation.* Ce temps comptabilisera seulement le temps de la recherche effective pendant l'observation. Les temps de présentation de la recherche et d'auto notation ne sont donc pas comptabilisés ;

- *Nombre de sites visités.* On comptabilise le nombre de fois où l'internaute clique sur un lien dans le moteur de recherche ou clique sur un lien dans un site renvoyant sur un autre site. Si le site a préalablement déjà été visité il est quand même comptabilisé ;

- *Nombre de pages supplémentaires visitées* (en plus de la première page). On comptabilise le nombre de fois où l'internaute clique sur un lien dans un site affichant une page du même site. Les pages réaffichées sont comptabilisées ;

- *Succès ou échec de la recherche.* Indique si l'information recherchée, document ou réponse à une question, a été trouvée. Quatre formes de réponses sont possibles : oui (succès total), non (échec complet) ou partiellement ; dans ce dernier cas, on notera s'il y a, malgré tout, peu d'éléments de satisfaction, ou au contraire, beaucoup d'éléments de satisfaction ;

- *Nombre de requêtes.* On comptabilise le nombre de fois où l'internaute envoie une requête au moteur de recherche. Les requêtes renvoyées sont comptabilisées.

Pour chaque requête, nous avons également sauvegardé :

- *Les mots-clés utilisés.* La requête est archivée à l'identique, les fautes d'orthographe et les caractères spéciaux sont conservés ;

- *Le nombre de mots-clés utilisés.* Chaque mot est comptabilisé. Les expressions entre guillemets sont comptabilisées comme un mot-clé ;

- *Le nombre de sites restitués.* On note le nombre de sites théoriques retournés par le moteur de recherche en réponse à une requête ;
- *Le caractère multi-langue des mots-clés utilisés.* Si la même requête contient des mots de langues différentes, elle est notée comme une requête multi-langue.

### 3.2. Les participants à l'étude

Prénom	Age	Usage	Expérience	Temps max. théorique	Situation
--------	-----	-------	------------	----------------------	-----------

#### – Les adultes

Jean	16 ans	Quotidien	> 5ans	15 minutes	Etudiant
Annie	72 ans	Mensuel	> 5ans	25 minutes	PDG retraitée
Georges	49 ans	Jamais	Nulle	15 minutes	Artisan
Marie	37 ans	Rare	< 1 an	15 minutes	Infirmière
Yasmina	20 ans	Mensuel	> 5ans	15 minutes	Employée

#### – Les enfants

Guillaume	11 ans	Hebdomadaire	Entre 1 et 5ans	15 minutes	Collégien
Céline	11 ans	Rare	< 1 an	15 minutes	Collégienne
Lou	7 ans	Hebdomadaire	Entre 1 et 5ans	15 minutes	Scolaire
Paul	7 ans	Rare	Jamais	15 minutes	Scolaire

NOTE. — Une présentation vidéo des internautes ayant participé à l'étude est consultable sur le site <http://sissiprojet.free.fr>. Nous avons essayé de choisir un échantillon représentatif, en âges, genres et niveaux sociaux.

### 3.3. Les observations

Les neuf utilisateurs ont effectué 21 recherches composées de 113 requêtes différentes, dont 15 requêtes multi-langues. Ces requêtes ont retourné au total 829 825 484 sites. L'analyse des observations effectuées lors des 21 recherches, en fonction des différents indicateurs présentés au paragraphe 2, a permis de dresser un ensemble de résultats présentés dans ce qui suit. Ces résultats sont analysés et discutés au paragraphe 3.5.

#### – Mesures sur les recherches :

	Temps consacré	Nombre de requêtes
Maximum	01h 03mn 30s	20
Minimum	00h 01mn 30s	0
Moyenne	00h 18mn 45s	4,5

**Tableau 1.** Les recherches – Temps et nombre de requêtes par recherche.

– Mesures sur les requêtes :

	Nombre de mots-clés	Nombre de sites restitués
Maximum	8	230000000
Minimum	1	0
Moyenne	3,44	7343588

**Tableau 2.** *Les requêtes – Nombre de mots-clés et nombre de sites restitués par requête.*

	Nombre de sites visités	Nombre de pages visitées dans ces sites	Nombre de pages des sites visités
Maximum	44	17	9
Minimum	1	1	0
Moyenne	8,73	3,53	2,6

**Tableau 3.** *La navigation – Nombre de sites visités (lus), nombre de pages visitées sur ces sites, nombre de pages des sites visités.*

Utilisateurs	Usage	Expérience	Pourcentage de requêtes avec réponse pertinente
Annie	rare	> 5 ans	40
Céline	mensuel	< 1 an	50
Georges	jamais	0	40
Guillaume	hebdomadaire	entre 1 et 5 ans	100
Jean	quotidien	entre 1 et 5 ans	60
Lou	jamais	entre 1 et 5 ans	50
Marie	rare	< 1 an	60
Paul	jamais	0	50
Yasmina	mensuel	> 5 ans	33
Moyenne	N/A	N/A	53,3

**Tableau 4.** *Recherches avec succès – Pourcentage moyen de requêtes avec réponses pertinentes.*

– Mesures sur les mots-clés :

Avec 3,44 mots-clés en moyenne par requête (la plupart des requêtes comprenant entre 2 et 6 mots-clés), nous pouvons noter que ce nombre est supérieur à celui obtenu par les observations précédentes effectuées sur le Web, y compris par celles rapportées dans la littérature (cf. paragraphe 2). Mais même avec une augmentation du nombre de mots-clés par requête, le nombre moyen de sites restitués est énorme (cf. figure 1 et 2 en annexe). Le nombre moyen de sites retournés, en utilisant principalement Google, est supérieur à 7 millions. Ainsi, nous pouvons affirmer que

les mots-clés utilisés dans l'observation ne sont pas assez efficaces et obligent toujours à filtrer manuellement la masse d'information restituée à partir du Web.

D'après les observations (cf. figure 2 en annexe), le nombre de sites retournés décroît avec l'augmentation du nombre de mots-clés, jusqu'à 6 mots-clés. On peut noter que le phénomène s'inverse au-delà de 6 mots-clés, les performances décroissent. Ceci implique que, conformément aux résultats déjà obtenus lors d'observations précédentes, augmenter le nombre de mots-clés de la requête ne suffit pas pour augmenter la précision des résultats. Il faudra trouver d'autres moyens.

### 3.4. Ressenti des utilisateurs

Avec une satisfaction générale moyenne de 6,25/10 (cf. tableau 5), le ressenti des utilisateurs sur Internet (en tant que base d'information) et sur la capacité des moteurs de recherche à (re)trouver l'information pourrait être amélioré. Quoi qu'il en soit, les résultats sont très différents d'un utilisateur à l'autre (cf. Tableau 5).

Utilisateurs	Lisibilité	Intérêt	Temps perçu	Note globale
Annie	7,4	6	7	6,3
Céline	7,5	6	4	4
Georges	5,5	5,5	6,5	6
Guillaume	7,5	7	7,5	7
Jean	8,4	6,5	5,8	6,3
Lou	2	3,5	5	5,5
Marie	4,8	5,4	5,2	5,7
Paul	7	10	5	10
Yasmina	7	5,75	7,25	6,5
<i>Maximum</i>	<i>10</i>	<i>10</i>	<i>10</i>	<i>10</i>
<i>Minimum</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>
<i>Moyenne</i>	<i>6,41</i>	<i>6,05</i>	<i>6</i>	<i>6,25</i>

**Tableau 5.** Notes moyennes de satisfaction pour chaque utilisateur (entre 1 et 10).

L'exercice a consisté à leur faire rechercher de l'information selon un protocole précis afin de pouvoir étudier leur comportement. Les utilisateurs ont tous joué le jeu et ont ainsi été très affectés par le fait de trouver une réponse ou de ne pas en trouver. Le temps réel passé sur la recherche ne révèle pas vraiment la note globale (cf. figures 3). Même la perception du temps passé ne cadre pas avec le temps réellement consacré (cf. figure 4). La note est plutôt à corrélérer avec la satisfaction de l'utilisateur de trouver une réponse ou la déception de ne pas trouver de réponse (cf. figure 5). En résumé, la perception du temps passé et l'intérêt pour un site semblent être plus en corrélation avec le succès ou l'échec de la recherche qu'avec tout autre chose (cf. figures 6 et 7). En effet, les notes fournies par les utilisateurs doivent être considérées avec prudence. Les utilisateurs étaient souvent obsédés par le fait de

trouver une réponse. Lorsqu'ils trouvent une réponse, le temps leur paraît court et la recherche intéressante, par contre, lorsqu'ils ne trouvent pas, le temps leur paraît long et la recherche ennuyeuse.

### **3.5. Difficultés rencontrées par des utilisateurs**

Les difficultés rencontrées par les utilisateurs se situent au niveau de chacune des tâches élémentaires qui composent la tâche globale de recherche d'information sur le Web. Ces difficultés sont principalement de quatre ordres : (1) trouver les mots-clés efficaces, (2) faire un choix dans une liste longue et hétérogène, (3) extraire de l'information et (4) gérer le temps (temps réel de recherche et perception).

#### *3.5.1. Difficultés pour trouver les mots-clés efficaces*

Avec un nombre moyen de sites Web trouvés par recherche à 7 millions, nous pouvons certifier que les 3,44 mots-clés employés dans les requêtes ne sont pas assez efficaces pour filtrer le Web. Les utilisateurs ont des difficultés pour employer plus de mots-clés et pour choisir ces mots, et lorsqu'ils le font cela devient contre-productif (taux de satisfaction décroît). Cet état de fait provient d'un problème de compétences des utilisateurs sur la nature même de la recherche: c.-à-d. que le plus souvent les utilisateurs ne connaissent pas assez le sujet pour enrichir la demande. Ceci provient également d'un mélange entre indolence et manque de confiance dans le moteur de recherche. D'autre part, l'utilisateur ne manie pas des mots dans une langue inconnue. Par exemple, pour rechercher de l'information sur la musique roumaine, personne n'a employé des mots-clés roumains (même si des outils de traduction étaient connus). Seulement *Jean* a utilisé des mots de la chanson anglaise à retrouver. *Jean* étant bilingue français-anglais, il a su mélanger le français et l'anglais. On peut alors remarquer que l'utilisateur choisit les mots-clés de sa requête en fonction de sa connaissance du sujet, et de sa connaissance du moteur de recherche utilisé. Il faut également considérer que les mots-clés peuvent être orthographiés de différentes façons. De fait, les jeunes utilisent de plus en plus le langage SMS. On a pu observer lors des expérimentations que les enfants trouvent (par erreur ?) l'information en formulant leur requête en langage SMS.

#### *3.5.2. Difficultés pour faire un choix dans une liste de sites restitués*

La liste des informations retournées par un moteur de recherches en réponse à une requête comprend un certain nombre d'éléments. Avec Google par exemple, comme le montre la figure 8, chaque item de la liste retournée intègre :

– le *titre du site*. Mais comme tous les sites n'ont pas un titre, dans ce cas, les moteurs, comme Google, utilisent d'autres métadonnées (metatags) telles que le titre de la page, l'auteur, l'URL,... pour construire une sorte de titre. De plus, lorsque ce titre est trop long, il est tronqué ;



– un *extrait* (« *snippet* »). Il s'agit en fait de bouts de phrases (quelques mots), entourant les mots-clés, extraits du site. L'extrait constitué en assemblant ces bouts de phrases n'a pas de sens globalement comme peut le montrer la figure 8 ;

– l' *URL* (*Uniform Resource Locator*) de la page ou du site.

On peut facilement imaginer un débutant facilement désabusé face à une telle liste. Le titre n'a pas forcément de sens, l'extrait est juste une liste de bouts de phrases et l'URL ressemble à du code binaire. Par exemple, *Annie*, 70 ans, a soumis à Google la requête suivante « Natura 2000 marais de Gabarret » où « Natura 2000 » est une organisation écologiste française et Gabarret est un petit village français. Elle a obtenu la réponse de la figure 8... Le titre est en anglais, l'extrait est en français et parle de la Tour Eiffel, l'URL est une adresse IP (Internet Protocol).

All the inside info that you need to know. It's right here waiting...  
Hauteur de la tour EIFFEL hauteur de la Tour EIFFEL en 2000 haut lyonnais hautparleurs .  
marais immobilier particulier ...  
64.41.125.45/cgi\_bin/extras/extra.pl?ref=215 23k - [En cache](#) - [Pages similaires](#)

**Figure 8.** Liste typique de sites restituée par Google

### 3.5.3. Difficultés pour extraire de l'information des objets trouvés

*A partir d'une page* : Il est difficile pour les utilisateurs d'obtenir l'information de la page. Tous les utilisateurs peuvent manquer des informations sur une page. Ceci peut provenir du fait que :

- ils cessent de lire la page avant que l'information « pertinente », ou du moins intéressante, ait été atteinte, la page est trop longue ;
- ils lisent juste la partie affichée de la page (effet fenêtre) ;
- la page est trop chargée, comme par exemple par des publicités ;
- ils utilisent une fonction de recherche sur un site au lieu d'utiliser un moteur de recherche sur le Web (Quid, Amazon...) ;
- ils utilisent toujours le même lien au lieu de lire la page (mode pleine page) ;
- ils cessent de lire si la page ou une partie de cette page est dans une langue inconnue.

*À partir d'un fichier* : l'information à extraire peut se trouver dans un fichier et non dans une page Web. Il faut alors savoir exploiter ce fichier, qu'il s'agisse d'une image (comme pour les partitions musicales par exemple), ou de tout autre type de fichier. Par exemple, *Jean* recherchait une partition de musique. Il a trouvé un fichier MIDI (son). Mais il n'a pas su qu'il était possible d'extraire la partition à partir d'un fichier MIDI ((Musical Instrument Digital Interface)) et abandonné déçu.

### 3.5.4. Difficultés pour gérer les temps de recherche et leur perception

La majorité des utilisateurs ont déclaré que le temps qu'ils ont passé sur une recherche ne dépasse pas 15 minutes. Or il se trouve que le temps moyen calculé, pour une recherche, est de 18 minutes et 45 secondes. De plus, 50% des recherches

ont largement dépassé les 15 minutes. Si le temps consacré à la recherche est moins important que le succès ou l'échec dans l'attribution de la note générale moyenne, au-delà de 15 minutes, l'appréciation fait systématiquement apparaître une certaine déception (seulement 5,5/10).

### **3.6. Conclusion sur les observations**

Les utilisateurs d'Internet sont déçus par le processus de recherche d'information. Ils ont des difficultés pour trouver les mots-clés efficaces et pour extraire l'information pertinente à partir de la liste restituée, comme à partir des sites eux-mêmes. Au regard des observations, des questions et des difficultés des internautes, il apparaît que :

- le choix des mots-clés dans un référentiel sémantique ne semble pas suffisant pour retourner un nombre raisonnable de sites et pour garantir un ordonnancement lié à ce champ sémantique ;
- les internautes ne savent pas formuler des requêtes dans des langues étrangères, et lorsque les informations retournées sont dans une langue qui leur est étrangère, ils se considèrent en situation d'échec ;
- les internautes ne comprennent pas forcément les textes retournés par les moteurs de recherche ;
- les formats des informations ou des fichiers retournés ne leur permettent pas d'extraire, ni d'exploiter l'information pertinente.

Le développement d'un système « intelligent » permettant de simplifier les services offerts par Internet, basé sur une meilleure connaissance de l'internaute, de ses goûts, de ses compétences et de ses ressources, pourrait contribuer à une meilleure adaptation des sites retournés (réponses). Par exemple, la traduction automatique vers des langues connues de l'internaute, l'indication du contenu et de la nature des sites par un système d'icônes serait de nature à rassurer. De même, la proposition de changement de format ou de nature de l'information par des services Web est également une solution possible pour aider tout un chacun à exploiter au mieux la connaissance du Web.

## **4. Une aide individuelle aux internautes par l'usage de services Web.**

### **4.1. Identifier les capacités individuelles (y compris les handicaps) et les connaissances des utilisateurs**

Aujourd'hui, des capacités, les goûts et les connaissances des utilisateurs ne sont pas pris en compte dans les processus de recherche d'information sur le Web. Les moteurs de recherche fonctionnent toujours de la même manière quel que soit l'utilisateur, alors que les utilisateurs s'avèrent être très différents les uns des autres.

L'objectif général de nos recherches vise à proposer des services Web permettant d'adapter la recherche d'information à un utilisateur. Le but est de l'aider dans le processus de recherche en l'intégrant au processus lui-même. Pour inclure l'utilisateur dans le processus, nous définissons un utilisateur comme un ensemble de ressources. Ces ressources, enregistrées dans un profil, peuvent être divisées dans deux parts classes :

- ressources liées à la communication. Ces ressources décrivent les capacités visuelles, auditives,... des utilisateurs, comme les média de restitution que nécessitent les capacités (ou incapacités) des utilisateurs ;
- ressources liées aux connaissances. Ces ressources décrivent les connaissances des utilisateurs telles que les langues connues (lues, parlées), les connaissances techniques sur Internet,...

Dans ce contexte, nous devons considérer le handicap au sens général du terme, et tous les handicaps. En réalité, face à Internet nous sommes tous plus ou moins « handicapés ». Le handicap n'est pas toujours là où nous pensons qu'il est. Pour rechercher sur le Web, une faible vue ou une faible audition peut être moins pénalisante que la non connaissance de langues étrangères.

Quoi qu'il en soit, le profil de l'utilisateur, pour pouvoir réellement l'aider dans ses recherches, doit intégrer toute information utile sur ses capacités et incapacités, ses goûts, ses expériences, les recherches effectuées et les mots-clés usuels,...

## **4.2. Offrir des services pour aider l'utilisateur à être plus efficace et moins déçu**

Nous pouvons proposer des solutions pour aider les utilisateurs et éliminer les difficultés. Ces solutions peuvent être individuelles, chaque utilisateur ayant besoin de solutions adaptatives. Ces solutions peuvent être très complexes à mettre en application. L'idée est de créer ou utiliser pour chaque solution un (des) service(s) Web indépendant(s). En partant des observations effectuées, nous pouvons déjà imaginer une liste de services susceptibles d'aider nos utilisateurs. Nous avons intégré et expérimenté certains de ces services dans l'architecture SISSI présentée au paragraphe suivant (4.3).

### **4.2.1. Des services pour aider nos utilisateurs et tout utilisateur**

Nos utilisateurs ont vécu des difficultés que nous avons identifiées (paragraphe 3.5). Il s'agit ici de trouver quels services pourraient être proposés en face de ces principaux obstacles. Il s'agit aussi de services généraux susceptibles d'aider tout utilisateur à rechercher les informations souhaitées, ou de l'information en général.

Quels services pour aider nos internautes dans leurs *Difficultés à trouver les mots-clés efficaces* ?

- *formuler ses requêtes* en reformulant ses requêtes initiales pour les adapter et les ramener dans un espace sémantique approprié au Web et à l'espace de recherche

visé par l'utilisateur. L'espace sémantique est déterminé par une adaptation de la connaissance, des goûts et préférences de l'utilisateur, en se basant en particulier sur ses précédentes recherches (espaces sémantiques déjà visités) ;

- *traduire des acronymes* (par exemple des sigles, des mots-clés SMS) en leur équivalent officiel correctement orthographié (et inversement) ;

- *traduire les requêtes et/ou les résultats dans une langue spécifique* soit dans des langues connus de l'internaute soit dans des langues liées à la recherche (cf. recherche de partition de l'hymne roumain) ;

- *Collaborer avec d'autres utilisateurs* par la mise en commun de requêtes et de ressources ;

- *corriger ou créer des fautes d'orthographe* de façon à élargir la recherche.

Quels services pour aider nos internautes dans leurs « Difficultés pour faire un choix dans une liste de sites restitués » ?

- *donner une lecture explicite des URL* ;

- *afficher un extrait plus explicite et plus détaillé* (autres que bouts de phrases) ;

- *placer une iconographie* précisant le contenu ou le type du site (web marchand, encyclopédie, jeux, forum ...) spécialement à l'usage des enfants.

Quels services pour aider nos internautes dans leurs « Difficultés pour extraire de l'information des objets trouvés » ?

- *mise en évidence des mots-clés* ou des paragraphes contenant ces mots-clés ;

- *retirer les objets polluants* (Images, pubs) ;

- *afficher une image* (plan) de la page pour aider à une localisation ;

- *extraire du contenu* (ou des extraits de contenu) à partir de fichiers spécifiques tels que fichiers images, audio, etc... pour le rendre exploitable sous un autre format. Par exemple, pour des musiciens, extraire des partitions de fichiers midi (sons).

Quels services pour aider nos internautes dans leurs « Difficultés pour gérer les temps de recherche et leur perception » ?

- *contrôler le temps de recherche* en sauvegardant les recherches et les statuts de ces recherches (afin de les analyser et les réutiliser) ;

- *analyser les sites restitués* pour aider des utilisateurs à modifier leur demande.

#### 4.2.2. *Des services pour aider des utilisateurs « handicapés »*

Ces services Web pourraient être utilisés pour aider des personnes ayant des problèmes d'accessibilité (aveugles, déficients visuels, daltoniens, difficultés de compréhension,...). Par exemple, nous pouvons imaginer des services Web agissant de façon transparente pour :

- *garantir une taille de police minimale* pour afficher les pages ;

- corriger des couleurs pour obtenir des pages plus lisibles (texte, fond, etc...) ;
- ajouter des liens vers des définitions pour les mots-clés détectés comme étant des « mots difficiles » ;
- retranscrire oralement le contenu des pages (restitution orale ou en braille par exemple pour les personnes aveugles) ; etc...

#### 4.2.3. Des services pour modifier les formats des informations

Les services Web peuvent également extraire, reconstruire, compléter l'information d'un document initial afin de rendre son contenu accessible à un utilisateur dans un contexte donné. Par exemple, ces services Web peuvent, toujours de façon quasiment transparente pour l'utilisateur :

- sélectionner le bon éditeur ou transformer le contenu pour que l'utilisateur puisse l'ouvrir avec son propre éditeur ;
- extraire des fichiers depuis n'importe quel type d'archive pour envoyer le contenu par courrier électronique ; etc...

#### 4.3. Définir une architecture ouverte pour accueillir et sélectionner des services Web — SISSI.

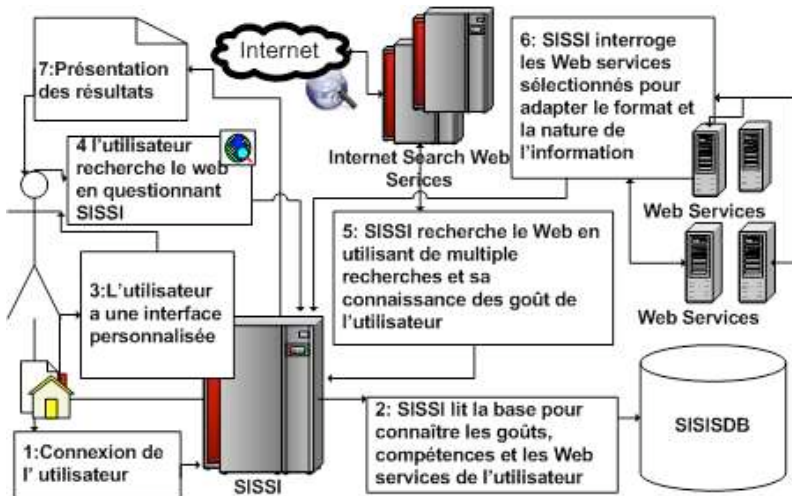


Figure 8. Architecture de transactions SISSI

##### 4.3.1. Les fondements de SISSI

Le but de nos recherches est de définir une architecture ouverte qui puisse accueillir et proposer des services Web à choisir. Les services Web pourront être

déposés par toute personne désirant mettre un service à la disposition de tous, l'architecture SISSI se chargeant ensuite d'orchestrer et de sélectionner le ou les services les mieux adaptés à la réalisation d'une tâche afin que l'utilisateur puisse se les approprier. Les services Web sont en charge d'aider l'utilisateur à rechercher de l'information et à lire l'information restituée, principalement, à la base, en adaptant la nature et le format des informations aux compétences des utilisateurs, à leurs connaissances et leurs goûts. L'architecture proposée est articulée autour d'un service central appelé SISSI: « Système Individualisé de Simplification des Services d'Internet » (cf. figure 8). Les définitions des services Web et les descriptions des utilisateurs sont stockées dans une base de données appelée SissiBD. Les services Web enregistrés dans la base SissiBD (annuaire UDDI — *Universal Description, Discovery and Integration*) présentent leurs services à l'utilisateur qui peut les sélectionner et les paramétrer, en fonction de ses besoins et de ses goûts.

#### 4.3.2. *SISSI et la communication entre services Web*

SISSI n'inclut aucun service Web. C'est un connecteur entre services Web répertoriés. SISSI peut être vu comme un client générique de services Web ou comme un service Web générique (une IHM générique par exemple).

La description des services Web est basée sur UDDI. Les services Web sont décrits avec plus de précisions dans la base SissiBD, de sorte à ce que les utilisateurs puissent comprendre les résultats qu'ils attendent, et que SISSI puisse communiquer avec eux.

La communication de SISSI est basée sur SOAP (*Simple Object Access Protocol*) et nous savons que SOAP est un protocole lourd. Mais l'information Web manipulée par SISSI concerne principalement des pages HTML ou du code, utilisé du côté client. Les objets complexes, tels que des images, restent sur le serveur HTTP. Si un service Web a besoin d'un de ces objets, il sera de sa responsabilité de le charger à partir du serveur initial de ressource.

SISSI inclus ici un processus de recherche. Un utilisateur pourra aussi invoquer les services de SISSI directement en lui fournissant un URL de document à traiter.

## 5. Conclusion

Les premières expérimentations et études de faisabilité menées avec SISSI indiquent une certaine complexité. Si le modèle d'architecture semble être utilisable, nous avons trouvé quelques nouveaux défis :

- le langage HTML, tel qu'il est utilisé dans les pratiques de tous les jours, n'est pas normalisé. Or, HTML se trouve être le principal support de restitution d'informations sur le Web ; aussi, agir sur le langage HTML, pour modifier la structure ou la présentation des documents, en tant que langage générique est aujourd'hui un travail nécessaire (bien que très ingrat) si on veut pouvoir

personnaliser et adapter l'information restituée. En raison de la complexité de gestion du langage HTML, écrire des services Web efficaces est très difficile ;

- SISSI est un service générique basé sur des services. Cela signifie que si un service s'écroule, SISSI tombe avec lui. Il nous faudra très rapidement nous intéresser à la tolérance aux fautes ;

- même s'ils emploient un pseudo, des utilisateurs ont le droit de se retirer d'une base où leurs compétences, connaissances, goûts et handicaps sont stockés, particulièrement si cette base enregistre leurs activités.

Cependant, conformément aux travaux connexes, aux observations et aux expérimentations menées avec SISSI, nous pouvons conclure que l'utilisation d'une architecture ouverte basée sur des services Web peut être une solution pour aider les internautes dans leur recherche d'informations. En effet, elle peut être un moyen d'aider les utilisateurs à faire face à certaines situations, plus particulièrement quand ils éprouvent quelques difficultés pour accéder à l'information recherchée ou à exploiter l'information trouvée pour des problèmes de forme ou de médium de restitution (internautes handicapés, utilisateurs non multilingues,...).

Pour conclure, les architectures basées sur des services comme celle que nous avons mise en application (SISSI) doivent être librement extensibles afin d'offrir toujours plus. A présent, nous devons expérimenter cette extensibilité en laissant n'importe quel utilisateur ajouter de nouveaux services quand il le juge nécessaire. De nouvelles observations pourront alors être faites pour analyser les capacités des internautes à interpréter les résultats fournis par l'utilisation de services cachés.

Enfin une meilleure connaissance des utilisateurs (goûts, compétences, contextes, ressources) permettra la mise en œuvre de communautés mutualisant les recherches, les résultats et les ressources. Ces informations pourront être utilisées à affiner les recherches par la création de contextes sémantiques associés.

## 6. Bibliographie

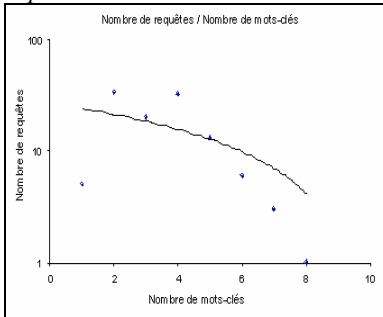
- Alexander P.A., Jetton T.L., Learning from text: A multidimensional and developmental perspective. In M.L. Kamil, P. Mosenthal, P.D. Pearson, and R. Barr (Eds.) *Handbook of reading research*, volume III, pp. 285-310, Mahwah, NJ: Erlbaum, 2002.
- Berkowitz B., Moving every child ahead: The Big6 success strategy. *Multimedia Schools*, 9, p. 17-22, 2002.
- Bransford J.D., Brown, A.L.O., Cocking R.R. (Eds.), *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press, 2000.
- Coiro J.L., Rethinking comprehension strategies to better prepare students for critically evaluating content on the Internet. *The NERA Journal*, 39(2), p. 29-34, 2003.
- Eagleton M.B., Guinee K., Langlais K., Teaching Internet literacy strategies: The hero inquiry project, *Voices from the Middle*, 10, p. 28-35, 2003.

- Gay L.R., Airasian P., *Educational research: Competencies for analysis and application*. Upper Saddle River, NJ: Prentice Hall, 2000.
- Guinee K., Eagleton M.B., Hall T.E., Adolescents' Internet search strategies: Drawing upon familiar cognitive paradigms when accessing electronic information sources, *Journal of Educational Computing Research*, 29, p. 363-374, 2003.
- Henry L.A., "Information search strategies on the Internet: A critical component of new literacies", *Webology*, 2(1), Article 9, 2005.
- International Reading Association, *Integrating literacy and technology in the curriculum: A position statement*. Retrieved April 22, 2005, from [http://www.reading.org/downloads/positions/ps1048\\_technology.pdf](http://www.reading.org/downloads/positions/ps1048_technology.pdf)
- Janessick V.J., The dance of qualitative research design: Metaphor, methodolarty, and meaning. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research*, p. 209-219, 1994. Thousand Oaks, CA: Sage Publications, Inc.
- Jansen B.J., Spink A., Saracevic T., Real life, real users, and real needs: a study and analysis of user queries on the Web. *Information Processing and Management*, p. 207-227, 2000.
- Laurie A.H., Information Search Strategies on the Internet: A Critical Component of New Literacies. University of Connecticut, Connecticut, U.S.A, 2005.
- Leu D.J. Jr., Our children's future: Changing the focus of literacy and literacy instruction. *The Reading Teacher*, 53, p. 424-431, 2000.
- Leu D.J. Jr., Kinzer C.K., Coiro J., Cammack D.W.,. *Toward a theory of new literacies emerging from the Internet and other communication technologies*. In R. Ruddell and N. Unrau (Eds.) *Theoretical models and processes of reading* (5th ed.). Newark, DE: International Reading Association, 2004.
- Nachmias R., Gilad A., Needle in a hyperstack: Searching for information on the World Wide Web. *Journal of Research on Technology in Education*, 34, p. 475-486, 2002.
- O'Sullivan M., Scott T., Ph.D Teaching Internet Information Literacy: A Critical Evaluation, <http://www.infoday.com/MMSchools/mar00/osullivan&scott.htm>, 2000.
- Pressley M., Afflerbach P., *Verbal protocols of reading: The nature of constructively responsible reading*. Hillsdale, NJ: Erlbaum, 1995.
- RAND Reading Study Group. *Reading for understanding: Towards and R&D program in reading comprehension*. 2002. Available Online: <http://www.rand.org/multi/achievementforall/reading/readreport.html>
- Salton G., The SMART Retrieval System. Englewood Cliffs, NJ: Prentice-Hall, 1970.
- Salton G., McGill M.J., *Introduction to Modern Information Retrieval*. New York: McGraw-Hill, 1983.
- Savoy J., Picard J., Retrieval effectiveness on the Web. *Information Processing and Management*, 37, p. 543-569, 2001.
- Sutherland-Smith W., Weaving the literacy Web: Changes in reading from page to screen. *The Reading Teacher*, 55, p. 662-669, 2002.
- Williams P, Nicholas D., *Testing the usability of information technology applications with learners with special educational needs* (SEN) <http://www.blackwell-synergy.com/doi/full/10.1111/j.1471-3802.2006.00057.x>, 2006.
- Yin R.K., *Case study research: Design and methods*. Sage: Newbury Park, CA, 1989.

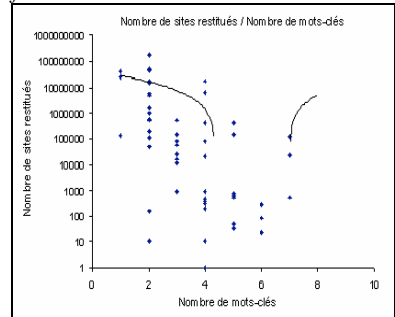


## 7. Annexe — Représentations graphiques des résultats des observations

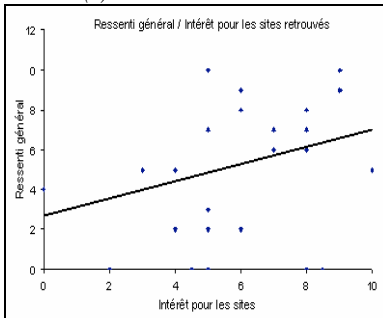
**Figure 1.** Répartition nombre de requêtes / nombre de mots-clés.



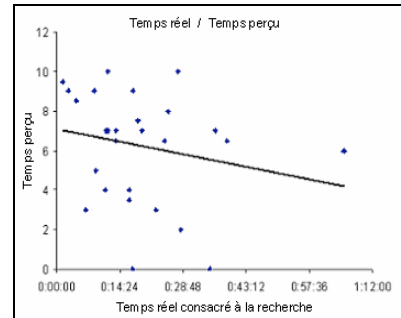
**Figure 2.** Nombre de sites restitués en fonction du nombre de mots-clés



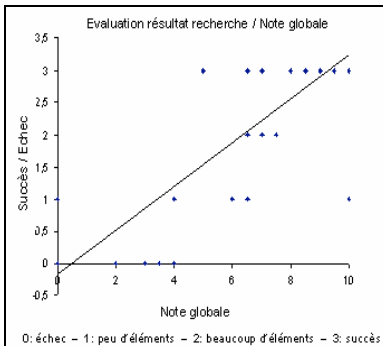
**Figure 3.** Ressenti général par rapport à l'intérêt subjectif pour le(s) site(s) retrouvé(s).



**Figure 4.** Temps réel consacré à la recherche par rapport au temps perçu..



**Figure 5.** Succès ou échec de la recherche par rapport à la note globale.



**Figure 6.** Succès ou échec de la recherche par rapport au temps subjectif consacré.

