

**Données e-comportementales et intention de comportement :
une méthode de collecte des données *log***

Résumé :

La littérature établit une relation entre le comportement sur un site en termes de nombre de pages vues, de temps passé par page ou de temps passé sur le site, les données *logs*, et l'intention de comportement de l'internaute. Cependant, les résultats demeurent hétérogènes et contradictoires. L'une des causes pourrait résider dans l'absence de prise en compte de variables modératrices ou médiatrices qui nécessiteraient plus de recherches. Cependant, le chercheur en marketing se heurte alors à un problème d'ordre méthodologique : la collecte de données *logs*, un problème d'autant plus invalidant quand il n'est pas administrateur et qui le restreint à l'utilisation de sites créés pour les besoins d'une expérience, ce qui réduit la validité même de ces expériences. Cet article de type exploratoire, à finalité méthodologique, permet de pallier à ce problème en proposant l'utilisation d'un outil de collecte de ces données non utilisé en marketing, le logiciel Netobserve.

Mots-clé : données logs, sites web, intention comportementale

E-consumer data and behavioral intention: a method to collect logs

Abstract:

The marketing literature states the relationship between the behavior on a website in terms of viewed pages, time spent on a page or time spent on the website, the logs, and the behavioral intention of the consumer. However, most of the time, results are heterogeneous and contradictory. One of the main reasons could be the influence of moderating or mediating variables which would require more studies. However, the marketing researcher has to cope with a methodological issue: collecting those logs, a problem all the more acute when he is not the web administrator, which often makes him use websites especially created for the purpose of an experiment and which reduces the validity of that experiment. This article which purpose is methodological helps to solve this problem in proposing a tool collecting those data, Netobserve.

Key-words: logs, websites, behavioral intention

Données e-comportementales et intention de comportement : une méthode de collecte des données *log*

Introduction

Dans un éditorial consacré au e-marketing, Dubois et Vernet (2001) encourageaient à « repenser la vision traditionnelle de la demande », tout en insistant sur l'aspect fortement concurrentiel de la toile. Dix ans plus tard, la concurrence est exacerbée. Attirer un internaute n'est pas chose facile et de surcroît, cela ne suffit pas: il faut ensuite le transformer en e-consommateur. Or, le taux d'abandon des paniers d'achat sur Internet est très élevé : on l'estime à 60 -75% avant la fin de la transaction alors même que le produit a été sélectionné et que la transaction est quasiment achevée (Rajamma, Paswan et Hossain, 2009). Qui est cet énigmatique e-consommateur qui veut acheter mais qui n'achète pas, ou qui ne revient pas ?

La recherche en marketing s'intéresse depuis une dizaine d'année à des données qui permettraient d'anticiper le comportement de l'internaute : c'est l'étude du parcours de navigation ou chemin parcouru sur un ou plusieurs sites par un consommateur. Ce parcours peut inclure des informations sur un même site telles que les pages visitées ou le temps passé sur chaque page ; il peut inclure également des informations sur les différents sites qui ont été visités et se traduit par des données laissées sur le serveur du site, les données *logs*.

Cependant, la littérature recense assez peu d'études qui établissent un lien entre les fichiers *logs* et l'intention de comportement, ce qui pourrait être en partie lié à la difficulté de collecte des données.

De prime abord, l'idéal semblerait être de collaborer avec une e-entreprise qui autoriserait l'accès aux données *logs* de son serveur : cependant, c'est un procédé qui demeure assez marginal du fait de la confidentialité de ce type de données (Hofacker et Murphy, 2009; Moe, 2003; Montgomery & al., 2004; Van den Poel et Buckinx, 2005). Par ailleurs, cette configuration expérimentale restreint les terrains envisageables: le chercheur n'aura pas nécessairement accès au site web qui pourrait convenir précisément à son étude de par sa finalité ou de par son design particulier. Il devra alors fonder son étude empirique avec comme terrain une entreprise par défaut qui lui aura permis l'accès au serveur de *logs*.

Cette difficulté méthodologique est le plus souvent résolue en administrant soi-même un site, ce qui rejoint la deuxième possibilité, qui est de créer en laboratoire un site web spécifique aux besoins d'une expérience (Ettis, 2008; Senecal, Kalczynski et Nantel, 2005). Cependant, la création de ce type de site pour les besoins d'une recherche nécessite beaucoup de temps ainsi que des compétences multidisciplinaires ; par ailleurs, la validité de ce type d'expérience serait limitée selon certains chercheurs (Diesbach, Galan et Chandon, 2006).

Dans cet article dont l'intérêt est essentiellement méthodologique, nous proposons une alternative : nous examinerons les possibilités offertes par un outil de collecte de *logs*, Netobserve. Ce logiciel permet la collecte de données quel que soit le site ; son utilisation ne nécessite pas d'avoir accès au serveur de l'entreprise et le logiciel permet le choix des données *logs* qui vont être étudiées.

A l'heure de Google analytics qui fournit au jour le jour un certain nombre de données logs traditionnelles comme le nombre de visites, le temps moyen passé sur le site ou le nombre de pages vues, notre recherche contribuera à établir le lien entre ces logs et l'intention

comportementale de l'internaute. Le manager de l'entreprise virtuelle pourra alors mesurer précisément l'influence de la modification d'une variable de l'interface ou plus globalement du parcours sur le site sur l'intention de comportement du consommateur.

Pour ce test, nous procéderons à un type de traitement classique de *logs* : l'examen du coefficient de Pearson ; notamment, nous étudierons le lien entre les données *logs* navigationnelles et temporelles et l'intention de comportement de l'internaute (intention d'achat, de retour et intention de recommander le site). Dans cet article, nous proposerons une méthodologie de traitement des données adaptée au logiciel Netobserve et nous mettrons en évidence le potentiel de l'outil et ses limites.

Ce sera ainsi l'occasion de présenter au chercheur en marketing la possibilité d'aller plus loin vers la compréhension du comportement de l'internaute en conduisant une recherche sur le site qu'il aura précisément souhaité utiliser comme terrain.

1. Revue de la littérature

La psychologie environnementale montre comment l'environnement physique influence les comportements d'approche et d'évitement (Mehrabian et Russell, 1980). Le modèle Stimulus – Organism – Response (S.O.R) décrit la façon dont les variables de l'environnement physique influencent l'état émotionnel des individus, ce qui résulte en des comportements d'approche ou d'évitement. Ce modèle a donné lieu à plusieurs recherches portant sur la relation entre l'atmosphère d'un lieu de vente et le comportement du consommateur (Baker, Levy et Grewal, 1992; Donovan et Rossiter, 1982) : c'est également le cadre théorique traditionnellement utilisé dans les études portant sur la relation entre les données *logs* et l'intention de comportement de l'internaute (e.g. Ahmed, 2004; Dailey, 2004; Eroglu,

Machleit et Davis, 2001; Eroglu, Machleit et Davis, 2003; Galan et Gonzales, 2001; Helme-Guizon, 2001; Lemoine, 2008; McKinney, 2004; Mummalaneni, 2005; Richard, 2005; Richard et Chandra, 2005; Rosen et Purinton, 2004; Sautter, Hyman et Lukošius, 2004; Wang, 2003).

Or, le S.O.R ne prend pas en compte les variables situationnelles qui pourtant nous semblent être un aspect fondamental lors de la navigation sur le web. C'est le cas du but utilitaire ou hédonique de l'internaute dont l'influence sur les données *logs* ainsi que sur l'intention de comportement a fait l'objet de précédentes études (Moe, 2003; Montgomery, Li, Srinivasan & al. Liechty, 2004; Wolfenbarger et Gilly, 2001). Ainsi, nous pensons que le paradigme Personne x Objet x Situation (Belk, 1974; Belk, 1975; Bloch et Richins, 1983; Punj et Stewart, 1983) est approprié à l'étude de la relation entre les données de *logs* et l'intention de comportement de l'internaute.

Il existerait un lien entre la complexité du parcours de navigation et le comportement de magasinage sur le web. Ce sont les données *logs* ou données *clickstream* qui rendent compte de ce parcours de navigation. Les données *logs* sont enregistrées sur l'ordinateur alors que l'internaute effectue sa session de navigation. On peut restreindre l'analyse de ces données à un seul site : on parle alors de *clickstream intra-site*.

Il y aurait un lien entre les données *logs* et le comportement de recherche sur le site (Hofacker et Murphy, 2009), le taux de micro-conversion (Lee & al., 2001), l'intention d'acheter (Van den Poel et Buckinx, 2005), ou l'intérêt dans l'achat en ligne (Bucklin et Sismeiro, 2003). Certains chercheurs ont établi la relation entre la prise d'informations avant achat et la complexité du parcours de navigation comme le nombre de pages vues (Senecal, Kalczynski

et Nantel, 2005). D'autres études suggèrent que les données *clickstream* peuvent prédire la réalisation d'une tâche caractéristique comme la recherche interne sur le site ou la finalisation de l'achat (Kalczynski, Senecal et Nantel, 2006). Certains pensent que le temps alloué pour visiter le site est représentatif de la possibilité d'acheter en ligne (Senecal, 2007; Sismeiro et Bucklin, 2004): comme le soulignent Lohse et al.(2000) « *the longer the amount of time spent online, the greater the chance of making a purchase online* ».

Moe (2003) suggère également une relation entre les données *logs* et l'orientation motivationnelle de l'internaute (plutôt hédonique ou utilitaire) : elle étudie notamment le lien entre les données *logs* et le but de l'internaute, et l'intention d'achat de ce dernier. Montgomery et al.(2004) poursuivent cette étude en orientant leur recherche vers une modélisation des données *logs* selon l'orientation motivationnelle de l'internaute, le but étant de prédire l'intention d'achat.

Le lien entre les données *logs*, l'intention de comportement de l'internaute ou son orientation motivationnelle semble établi, cependant, le résultat de ces études demeure contradictoire, notamment en ce qui concerne la relation entre les données *logs* et l'intention d'achat. On peut supposer l'influence de variables modératrices ou médiatrices qui n'ont pas été clairement identifiées et qui sont la cause de telles divergences dans les résultats. On peut également supposer que certaines données *logs*, non explorées, seraient plus fiables que d'autres quand il s'agit de prédire le comportement de l'internaute.

En conclusion, le chantier qui met en scène la relation entre les données *logs* et l'intention de comportement semble assez loin d'être achevé, ce qui est en partie lié au problème de la collecte de données. Dans cet article, nous proposons de suppléer à ce problème grâce à un

outil qui rendrait son autonomie au chercheur souhaitant étudier ce type de données, et qui n'est pas utilisé dans la recherche en marketing, le logiciel Netobserve.

2. Méthodologie

Le site que nous proposons pour cette étude au demeurant exploratoire est celui de l'entreprise Nature et Découvertes¹ : c'est un site complet, à la fois marchand et informationnel. Il caractérise le type de site qui serait difficile à recréer en laboratoire du fait de la densité de l'information et du nombre de pages que l'internaute peut visiter ; il autorise des sessions de navigation potentiellement longues et complexes ce qui peut rendre l'automatisation du traitement des *logs* d'autant plus intéressante.

Notre échantillon est constitué de 150 internautes, étudiants, enseignants ou personnel administratif de l'ESC Bretagne Brest, ainsi que des membres de l'entreprise Produits en Bretagne. Il s'agit d'un échantillon de convenance puisque les testeurs ont répondu favorablement à un mail leur proposant de participer à une recherche portant sur l'étude d'une session de navigation. Les testeurs pouvaient naviguer librement sans restriction temporelle, puis ils devaient remplir un questionnaire papier à la fin de leur session de navigation permettant de mesurer leur intention de comportement (intention de retour, d'achat et intention de recommander le site web).

La mesure de l'intention comportementale est une traduction de l'échelle de Kaltcheva et Weitz (2006) qui démontre des résultats psychométriques satisfaisants avec α allant de .884 à .907 selon les études réalisées. C'est une échelle de Likert en 9 points, allant de 1= pas du tout d'accord à 9=tout à fait d'accord (tableau 1).

¹ natureetdecouvertes.com

Échelle de mesure de l'intention de comportement traduite de l'échelle de Kaltcheva et Weitz (2006)
J'aimerais retourner sur ce site web
Je serais prêt à acheter des produits/services sur ce site web
Je serais d'accord pour recommander ce site web à mes amis

Tableau 1: mesure de l'intention de comportement

Netobserve de la société ExploreAnywhere Software, LLC² est une administration à distance qui permet de se relier à distance à un PC (à l'intérieur ou en dehors d'un réseau). On pourrait placer ce logiciel dans la catégorie des *spyware*, bien que son usage soit exclusivement lié à des fins de surveillance familiale ou d'environnement d'affaire et aucunement à un piratage.

Le logiciel permet d'obtenir plusieurs types de données comme celles relatives à la navigation sur le web ou à l'activité Windows ; c'est également un *keylogger* c'est-à-dire que le logiciel conserve en mémoire l'ensemble de l'activité de frappe clavier. Des captures d'écrans peuvent être réalisées selon une cadence que l'administrateur définit à l'avance. Il est également possible de prendre des clichés de l'utilisateur final à intervalles prédéterminés.

L'utilisation de ce logiciel présente des avantages indéniables, mais également des inconvénients qu'il convient de prendre en considération. L'avantage majeur réside dans la facilité de la collecte qui ne nécessite pas l'accès au serveur de *logs* du site. Par ailleurs, il convient de noter la façon non intrusive d'obtenir ces données. Le deuxième avantage notable est celui de la fiabilité du logiciel qui a été éprouvée dans une précédente étude. Il en est de même pour sa souplesse d'utilisation ou la richesse des données récoltées même si dans le

² <http://www.exploreanywhere.com/>

cadre précis de cette recherche nous ne nous intéressons qu'à un seul type de données : les *logs* relatifs à l'activité web.

Les limites du logiciel sont essentiellement techniques : Netobserve fonctionne sous système d'exploitation Win95, Win98, WinME, WinNT 4.x, Windows2000 ou WinXP ; il n'existe pas de version compatible Windows Vista ou Windows 7. Avec ces deux derniers systèmes d'exploitation, il faut utiliser une machine virtuelle telle que Virtual PC³ qui permet de faire fonctionner en même temps plusieurs systèmes d'exploitation sur une même machine physique.

Une deuxième limite consiste à devoir récolter les *logs* à la fin de chaque session de navigation, puis d'effacer l'ensemble des données avant la session suivante. En effet, il convient de détourner les fonctionnalités premières de Netobserve qui n'est pas conçu pour la récolte de données différenciées multiutilisateurs, mais qui a été pensé pour enregistrer l'activité totale d'un poste client sur une période *t*. Après enregistrement d'une session, il faut donc sélectionner les données pertinentes à l'analyse en lançant simplement un navigateur web et en inscrivant dans la barre de navigation l'adresse IP de la machine sur laquelle on désire récolter les données.

2.1 Analyse des *logs*

L'analyse des données se fait en deux temps : une première exportation des données sous Excel et l'utilisation d'algorithmes vont permettre dans un deuxième temps l'utilisation d'un logiciel de statistiques. Les *logs* bruts obtenus avec le logiciel Netobserve (annexe 1) ne sont pas utilisables en l'état : ils doivent être exportés sous un tableur comme Excel pour être

³ <http://www.microsoft.com/windows/virtual-pc/>

traités avec une série de formules logiques du type SI _ ALORS, ce qui fournit des données temporelles (temps passé par page et temps passé sur le site) et navigationnelles (nombre de pages visitées). Sur un échantillon plus large, le chercheur peut concevoir des requêtes SQL qui permettront un accès immédiat aux données dont il a besoin pour procéder à son analyse.

Le premier traitement sous Excel a permis de caractériser l'échantillon des testeurs et d'obtenir une vue d'ensemble des sessions de navigation. Ainsi, la proportion des internautes dont le nombre de pages vues est supérieur à la moyenne ou dont le temps passé sur le site ou par page est supérieur à la moyenne est approximativement le même : il se situe entre 35 et 37 % du total soit près du tiers de la population des testeurs (annexe 2). En observant plus précisément la répartition des pages vues par tranches temporelles (figure 1), on remarque que la moitié des pages ont été scannées rapidement par le regard (50 % des pages vues en moins de 20 secondes).

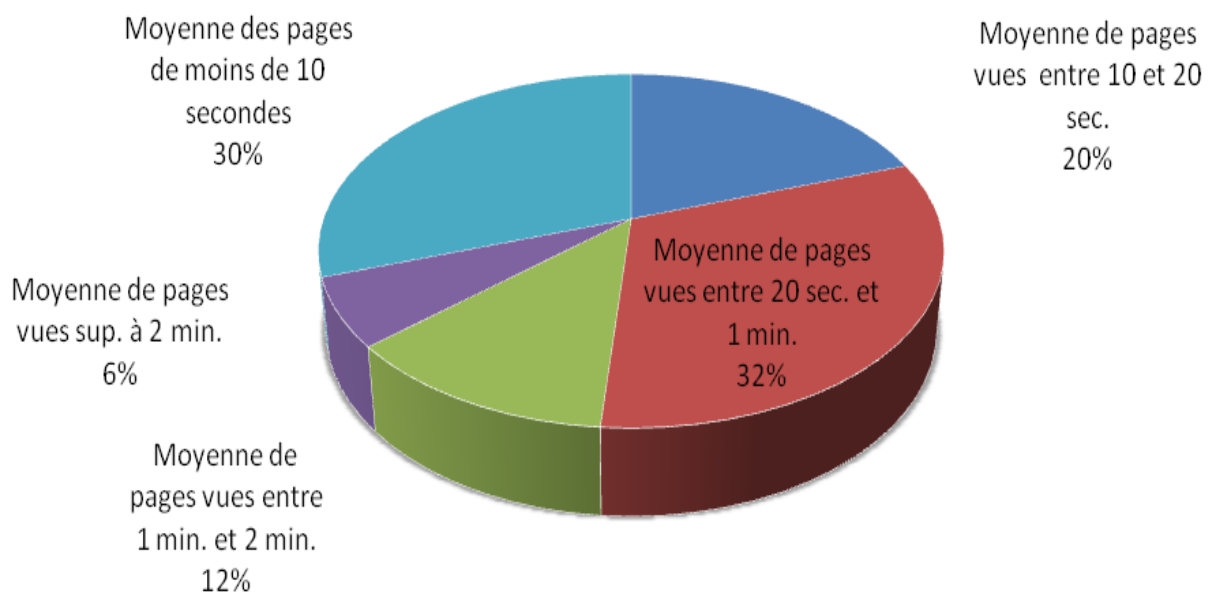


Figure 1: répartition des pages vues par tranches temporelles

Peu de pages ont ainsi été réellement observées (temps supérieur à 1 minute, ce qui correspond à 18 %). Deux internautes sur trois ont eu une session de navigation plutôt rapide que ce soit d'un point de vue temporel ou selon le nombre de pages visitées. Le temps passé par page, qui est assez faible, pourrait indiquer une lecture très superficielle voire un scan de la page.

2.2 Logs et intention de comportement

On examine la première source de données *logs* : le temps passé sur le site. Notamment, on s'intéresse à la relation entre le temps passé sur le site et l'intention comportementale de l'internaute : intention d'achat, de retour sur le site ou intention de recommander le site. Le temps passé sur le site le moins élevé est proche d'une minute alors que le maximum est de plus d'une heure (figure 2). La moyenne est de l'ordre de 15 minutes mais l'écart type important montre que les observations sont situées assez loin de la moyenne.

	N	Intervalle	Minimum	Maximum	Moyenne	Ecart type	Variance	Asymétrie		Kurtosis	
	Statistique	Statistique	Statistique	Statistique	Statistique	Statistique	Statistique	Statistique	Erreur std	Statistique	Erreur std
tps_site	150	01:04:10	00:01:10	01:05:20	00:15:51	00:11:48,774	502360,117	1,324	,198	1,791	,394
nb_pg_vu	150	81	2	83	22,84	17,197	295,746	1,324	,198	1,554	,394
tp_my_pg	150	00:03:15	00:00:02	00:03:17	00:00:49	00:00:29,068	844,952	1,789	,198	4,644	,394
pg10_20s	150	25	0	25	4,48	4,605	21,204	1,836	,198	3,767	,394
pg20s_1m	150	28	0	28	7,23	5,394	29,093	1,048	,198	1,010	,394
pg1m_2m	150	12	0	12	2,80	2,285	5,221	1,074	,198	1,251	,394
pg_sup2m	150	6	0	6	1,43	1,287	1,657	,963	,198	,871	,394
N valide (listwise)	150										

Figure 2: statistiques descriptives des données *logs* temps passé sur le site

L'examen des ratios « coefficient d'asymétrie sur erreur type » (non compris entre -2 et + 2) et « coefficient d'aplatissement sur erreur type » (non compris entre -2 et + 2) montre que la courbe ne suit pas une loi normale et si on observe le graphe de répartition gaussien, on constate un certain nombre de valeurs aberrantes après suppression desquelles le temps passé sur le site suit une loi normale, et ce, pour 73 internautes comme en témoigne l'étude des

ratios d'asymétrie et d'aplatissement ainsi que le test de Shapiro-Wilk (annexe 3). Étant donné que l'intention d'achat, de retour, ou de recommander le site suit également une loi normale pour cette population de testeurs (annexe 4), on procède à l'étude du coefficient de corrélation (annexe 5) qui ne permet cependant pas d'établir de relation significative entre le temps passé sur le site et l'intention de comportement.

On procède ensuite à l'étude de la deuxième source de données *logs*, le temps passé par page en cherchant une éventuelle relation entre le temps passé par page et l'intention de comportement de l'internaute. Le temps passé par page le moins élevé est de l'ordre de deux secondes alors que le maximum est de plus de trois minutes si on prend le groupe des 150 internautes (figure 2). L'examen de l'écart type témoigne d'observations assez éloignées de la moyenne. De même que pour le temps passé sur le site, la courbe du temps moyen passé par page ne suit pas une loi normale et après suppression des valeurs extrêmes et examen des tests de normalité, on procède à l'examen du coefficient de corrélation (annexe 6) qui ne montre pas de lien significatif entre les variables testées.

Enfin, on procède à l'étude de la troisième et dernière source de données *logs*, le nombre de pages vues. Notamment, on examine la relation entre le nombre de pages vues, qui varie entre 2 et 83 pages (figure 2) et l'intention de comportement de l'internaute. Comme dans les deux cas précédents, la courbe ne suit pas une loi normale ce qui nous amène à supprimer les valeurs extrêmes. Le calcul du coefficient de corrélation (annexe 7) ne montre pas de relation significative entre le nombre de pages vues et l'intention de comportement.

3. Limites et voies de recherche

Notre expérimentation fut confrontée à un certain nombre de limites qui peuvent en partie expliquer nos résultats. Tout d'abord, les tests ont été réalisés en conditions de laboratoire ; on

peut supposer que l'accès aux données logs dans des conditions naturelles de navigation aurait pu apporter des résultats différents. En effet, comme le dit ce testeur, « je ne sais pas si je serais restée si longtemps dans la vraie vie. Toute seule,...(silence) je sais pas. Peut être que j'ai eu ici le souci de trop bien faire ». Si on se place dans une situation marchande réelle, en se référant par exemple à un compte Google Analytics, on se rend compte que de nombreuses visites sont limitées à une seule page puisque les internautes aboutissent le plus souvent sur un site via des mots-clés et un moteur de recherche. La page de démarrage déterminant le but et l'orientation du site, l'internaute visite rarement plus d'une page quand mots-clés et but du site ne sont pas en adéquation avec ce qu'il anticipait. Or, dans notre étude, aucun des 150 internautes n'a visité une seule page. On pourrait en dire autant du temps passé sur le site ou du temps moyen passé par page. La méthodologie choisie qui imposait notre présence (enregistrement des *logs* et remise à zéro du compteur) a très certainement biaisé les résultats.

Par ailleurs, nous avons fait le choix de trois types de données log en fonction de la littérature et également pour des raisons techniques puisque ce sont les données les plus facilement obtenues avec le logiciel Netobserve. On peut se demander si d'autres types de données auraient pu apporter des résultats plus conformes à nos hypothèses, comme le taux de micro-conversion ou le nombre de pages produits visitées.

Enfin, si nous sommes partis d'une base de 150 testeurs, le traitement statistique adopté impose la normalité des données ce qui nous a conduit à éliminer les données extrêmes et donc à restreindre considérablement la taille de l'échantillon. De fait, la portée des résultats s'en trouve limitée. Peu après l'expérimentation, le site a été entièrement remanié, tant au niveau de l'ergonomie des pages qu'au niveau de la technologie utilisée puisqu'aujourd'hui le site Nature et Découvertes est en grande partie en Flash. Nous n'avons donc pas pu conduire

une nouvelle collecte de données. Il s'agit d'une contrainte majeure quand le chercheur souhaite travailler sur un site qui n'est pas créé pour les besoins d'une expérience : il ne contrôle ni l'ensemble des paramètres ni leurs modifications. Le chercheur souhaitant poursuivre l'étude des données logs devra s'interroger sur cette limite majeure. Dans certains cas, nous pensons qu'il peut être pertinent de repenser l'intérêt d'utiliser un site créé pour les besoins de la recherche.

Les limites de cette recherche constituent autant de voies de recherche, notamment méthodologiques. Tout d'abord, nous recommandons de ne pas travailler en monoposte : nous proposons une étude simultanée multipostes par groupes de marguerites (une marguerite étant constituée de quatre à cinq postes, tous équipés du logiciel Netobserve) dans un laboratoire multimedia. L'administration à distance du logiciel Netobserve permettra alors au chercheur la collecte simultanée de l'ensemble des données logs depuis un poste distant. Surtout, la collecte deviendra moins intrusive ; nous pensons que la sensation d'être observé sera alors plus faible et de fait, que la navigation des testeurs deviendra plus naturelle.

Conclusions

Les données *logs* qui sont les traces laissées sur le serveur d'un site web traduisent le comportement de l'internaute sur ce site lors d'une session de navigation. Or, la littérature montre que l'étude des *logs* permettrait d'établir un lien avec l'intention de comportement du consommateur, un intérêt certain pour le gestionnaire de l'entreprise qui n'utilise pas pleinement le potentiel de ces données.

Or, assez peu d'études sont recensées en marketing, avec des résultats qui restent contradictoires ; il conviendrait d'accentuer la recherche dans ce domaine si ce n'est que le

chercheur se heurte à la difficulté de la collecte des *logs*. Cet article a répondu à ce manque, en proposant l'utilisation d'un logiciel qui autorise l'accès à ces données quelque soit le type de site choisi pour l'expérimentation.

Dans cet article, nous proposons une méthodologie liée à l'utilisation de ce logiciel, lors d'une étude au demeurant exploratoire : l'étude de la relation entre l'intention de comportement (achat, retour, recommander le site), les données *logs* temporelles (temps passé sur le site, temps passé par page) et navigationnelle (nombre de pages visitées).

Concernant les résultats de l'étude à proprement parler, le temps passé sur le site, le temps passé par page et le nombre de pages visitées ne suivent pas une loi normale pour la population des 150 testeurs : nous avons éliminé les valeurs aberrantes ce qui a permis d'examiner le coefficient Pearson. Pour ces populations réduites, il ne semble pas y avoir de relation significative entre l'intention de comportement et les données *logs* : que ce soit le temps passé sur le site, le temps passé par page ou le nombre de pages visitées.

Les résultats ne sont pas ceux qui étaient escomptés et qui consistaient à établir la relation entre les données *logs* temporelles et navigationnelles et l'intention de comportement, ce qui peut être la conséquence de biais méthodologiques. Tout d'abord, si le logiciel proposé dans cette recherche permet de supprimer la contrainte du site créé pour les besoins de l'expérience lors de la collecte de données *logs*, il n'en demeure pas moins que les conditions de laboratoire constituent un biais important pour la validité des résultats obtenus et ce qui constitue la limite la plus importante de cette recherche. Lors d'une étude ultérieure, il conviendrait d'étendre la population test et de ne pas la limiter à un échantillon de convenance de façon à supprimer les biais d'échantillonnage.

Cependant, cette étude ouvre la voie vers de nouvelles recherches notamment en termes de modélisation, de par la facilité de collecte et de traitement des *logs*. On peut également envisager l'effet modérateur ou médiateur jusqu'à présent non exploré de certaines variables.

Bibliographie

- Baker J., Levy M. et Grewal D. (1992), An experimental approach to making retail store environmental decisions, *Journal of Retailing*, 68, 4, 445.
- Belk R. (1974), An exploratory assessment of situational effects in buyer behavior, *Journal of marketing research*, 11, 2, 156-163.
- Bloch P. H. et Richins M. L. (1983), Shopping without purchase: an investigation of consumer browsing behavior, *Advances in Consumer Research*, 10, 389-393.
- Bucklin R. E. et Sismeiro C. (2003), A Model of Web Site Browsing Behavior Estimated on Clickstream Data, *Journal of Marketing Research*, 40, 3, 249-267.
- Dailey L. (2004), Navigational web atmospherics explaining the influence of restrictive navigation cues, *Journal of Business Research*, 57, 7, 795-803.
- Diesbach B. P., Galan J. P. et Chandon J. L. (2006), Impact de la présence d'un agent virtuel incarné sur le pouvoir de rétention du site web : une analyse comportementale par les fichiers log., *11e JRMB*, Dijon,
- Donovan R. et Rossiter J. (1982), Store atmosphere: an environmental psychology approach, *Journal of Retailing*, 58, 1, 34-57.
- Dubois B. et Vernet E. (2001), Éditorial, *Recherche et Applications en Marketing*, Volume 16, No 3, 1.
- Eroglu S., Machleit K. A. et Davis L. M. (2001), Atmospheric qualities of online retailing: a conceptual model and implications, *Journal of Business Research*, 54, 2, 177-184.
- Eroglu S. A., Machleit K. A. et Davis L. M. (2003), Empirical testing of a model of online store atmospherics and shopper responses, *Psychology and Marketing*, 20, 2, 139-150.
- Helme-Guizon A. (2001), Le comportement du consommateur sur un site marchand est-il fondamentalement différent de son comportement en magasin : Proposition d'un cadre d'appréhension de ses spécificités, *Recherche et Applications en Marketing*, 16, 3, 25-38.
- Hofacker C. et Murphy J. (2009), Consumer web page search, clicking behavior and reaction time, *Direct Marketing: An International Journal*, 3, 2, 88-96.

- Kalczynski P., Senecal S. et Nantel J. (2006), Predicting on-line task completion with clickstream complexity measures: A graph-based approach, *International Journal of Electronic Commerce*, 10, 3, 121-141.
- Kaltcheva V. et Weitz B. (2006), When Should a Retailer Create an Exciting Store Environment?, *Journal of Marketing*, 70, 107-118.
- Lee J., Podlaseck M., Schonberg E. et Hoch R. (2001), Visualization and Analysis of Clickstream Data of Online Stores for Understanding Web Merchandising, *Data Mining and Knowledge Discovery*, 5, 1-2, 59-84.
- Lemoine J.-F. (2008), Atmosphère des sites web marchands et réactions des internautes, *Revue Française du marketing*, 217, 2/5, 45-61.
- Lohse G., Bellman S. et Johnson E. (2000), Consumer buying behavior on the Internet: Findings from panel data, *Journal of Interactive Marketing*, 14, 1, 15-29.
- McKinney L. N. (2004), Creating a satisfying Internet shopping experience via atmospheric variables, *International Journal of Consumer Studies*, 28, 3, 268-283.
- Mehrabian A. et Russell J. A. (1980), *An Approach to Environmental Psychology*, Cambridge, MA, The MIT Press.
- Moe W. (2003), Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream, *Journal of Consumer Psychology*, 29-39.
- Moe W. W. (2003), Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream, *Journal of Consumer Psychology*, 13, 1-2, 29-39.
- Montgomery A., Li S., Srinivasan K. et Liechty J. (2004), Modeling online browsing and path analysis using clickstream data, *Marketing Science*, 23, 4, 579-595.
- Mummalaneni V. (2005), An empirical investigation of web site characteristics, consumer emotional states and on-line shopping behavior, *Journal of Business Research*, 58, 4, 526-532.
- Punj G. et Stewart D. (1983), An Interaction Framework of Consumer Decision Making, *Journal of Consumer Research*, 10, 2, 181-196.
- Rajamma R. K., Paswan A. K. et Hossain M. M. (2009), Why do shoppers abandon shopping cart? Perceived waiting time, risk, and transaction inconvenience, *Journal of Product & Brand Management*, 18, 3, 188-197.
- Richard M.-O. (2005), Modeling the impact of internet atmospherics on surfer behavior, *Journal of Business Research*, 58, 12, 1632-1642.

Richard M.-O. et Chandra R. (2005), A model of consumer web navigational behavior: conceptual development and application, *Journal of Business Research*, 58, 8, 1019-1029.

Rosen D. et Purinton E. (2004), Website design: viewing the web as a cognitive landscape, *Journal of Business Research*, 57, 787-794.

Sautter P., Hyman M. R. et Lukošius V. (2004), E-tail atmospherics: a critique of the literature and model extension, *Journal of Electronic Commerce Research*, 5, 1, 14-24.

Senecal S. (2007), Clickstream et modélisation, *Atelier thématique du GERAD: modèles d'aide à la décision en marketing*, HEC Montréal,

Senecal S., Kalczynski P. et Nantel J. (2005), Consumers' Decision-Making Process and Their Online Shopping Behavior: A Clickstream Analysis, *Journal of Business Research*, 58, 11, 1599-1608.

Sismeiro C. et Bucklin R. (2004), Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach, *Journal of Marketing Research*, 41, 3, 306-323.

Van den Poel D. et Buckinx W. (2005), Predicting online-purchasing behaviour, *European Journal of Operational Research*, 166, 2, 557-575.

Wolfenbarger M. et Gilly M. (2001), Shopping online for freedom, Control and Fun, *California Management Review*, 43, 2,

Annexes

Données *logs* telles que représentées sur l'interface web du logiciel Netobserve

Remote Machine IP Address: 127.0.0.1 Remote Machine Name: netobserve

Website Activity Log

The following list contains a list of all websites that have been visited on the remote machine...

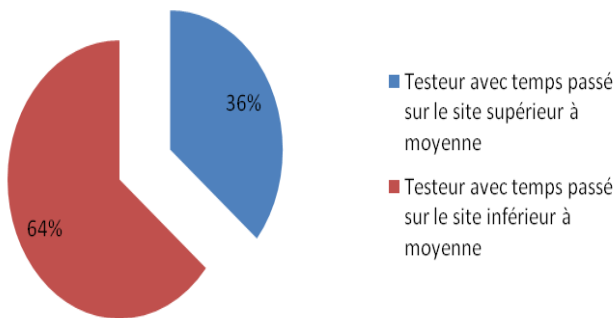
Username	Time Visited	Website
esc	mar. 04/27/10 @ 01:39:24 PM	NETObserve Control Panel
esc	mar. 04/27/10 @ 01:39:18 PM	NETObserve Control Panel Login
esc	mar. 04/27/10 @ 01:39:09 PM	NETObserve Control Panel
esc	mar. 04/27/10 @ 01:38:58 PM	NETObserve Control Panel Login
esc	mar. 04/27/10 @ 01:36:45 PM	Erreur de syntaxe
esc	mar. 04/27/10 @ 01:36:42 PM	http://
esc	mar. 04/27/10 @ 01:35:27 PM	Abc de la taille
esc	mar. 04/27/10 @ 01:35:13 PM	Libellule solaire dansante
esc	mar. 04/27/10 @ 01:34:27 PM	Nature & Découvertes : et si on vivait autrement ?

Données *logs* exportées sous Exel et traitées avec un algorithme

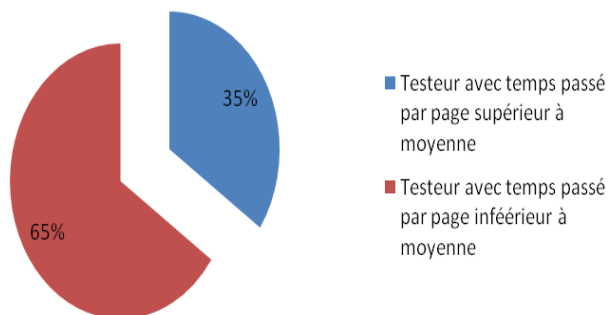
Pages visitées	temps passé par page	heure, minutes, secondes	Website Title	Website URL
1		11:24:13	Nature & Découvertes : et si on vivait autrement ?	http://www.natu
2	00:00:49	11:23:24	Magasin Nature & Découvertes en ligne. Effectuez vos achats nature : idées cadeaux, livres, obj	http://www.natu
3	00:00:24	11:23:00	Accueil	http://www.natu
4	00:00:06	11:22:54	Nature & Découvertes : et si on vivait autrement ?	http://www.natu
5	00:00:24	11:22:30	Accueil	http://www.natu

Annexe 1: traitement des données *logs*

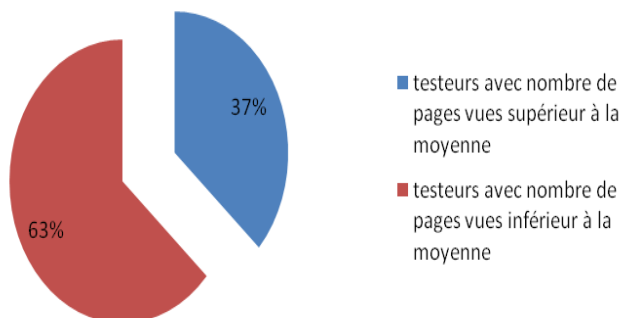
Temps passé sur le site



Temps moyen passé par page



Nombre de pages vues



Annexe 2: testeurs répartis selon le temps passé sur le site, le temps passé par page et le nombre de pages vues

		Statistique	Erreur standard	
tp_my_pg	Moyenne	00:00:57	00:00:01,362	
	Intervalle de confiance à 95% pour la moyenne	Borne inférieure	00:00:54	
		Borne supérieure	00:00:59	
	Moyenne tronquée à 5%	00:00:57		
	Médiane	00:00:56		
	Variance	74,182		
	Ecart-type	00:00:08,613		
	Minimum	00:00:45		
	Maximum	00:01:13		
	Intervalle	00:00:28		
	Intervalle interquartile	00:00:16		
	Asymétrie	,272	,374	
	Aplatissement	-1,199	,733	

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistique	ddl	Signification	Statistique	ddl	Signification
tp_my_pg	,124	40	,126	,933	40	,020

a. Correction de signification de Lilliefors

Annexe 3: Temps moyen par page. Tests de normalité après élimination des données aberrantes

ACHAT	Statistique	Erreur standard	ratios
asymétrie	-,436	,281	-1,5500536
aplatissement	-1,146	,555	-2,0637745
RETOUR			
asymétrie	-,506	,281	-1,7987616
aplatissement	-,593	,555	-1,0687246
RECOMM			
asymétrie	-,496	,281	-1,7636307
aplatissement	-,692	,555	-1,2461246

Annexe 4: temps passé sur le site. Normalité des courbes intentions d'achat, de retour et de recommander le site pour la population des 73 testeurs

Corrélations

		tps_site	achat	retour_s	recomman
tps_site	Corrélation de Pearson	1	,078	,095	,045
	Sig. (bilatérale)		,512	,422	,707
	N	73	73	73	73
achat	Corrélation de Pearson	,078	1	,617**	,599**
	Sig. (bilatérale)	,512		,000	,000
	N	73	73	73	73
retour_s	Corrélation de Pearson	,095	,617**	1	,706**
	Sig. (bilatérale)	,422	,000		,000
	N	73	73	73	73
recomman	Corrélation de Pearson	,045	,599**	,706**	1
	Sig. (bilatérale)	,707	,000	,000	
	N	73	73	73	73

** . La corrélation est significative au niveau 0.01 (bilatéral).

Annexe 5: étude du coefficient de corrélation pour le temps passé sur site et l'intention de comportement

Corrélations

			tp_my_pg	achat	retour_s	Recomman
tp_my_pg	Corrélation de		1	-,006	-,066	-,182
	Pearson					
	Sig. (bilatérale)			,970	,684	,262
	N		40	40	40	40
achat	Corrélation de		-,006	1	,679**	,544**
	Pearson					
	Sig. (bilatérale)		,970	,000	,000	,000
	N		40	40	40	40
retour_s	Corrélation de		-,066	,679**	1	,644**
	Pearson					
	Sig. (bilatérale)		,684	,000	,000	,000
	N		40	40	40	40
recomman	Corrélation de		-,182	,544**	,644**	1
	Pearson					
	Sig. (bilatérale)		,262	,000	,000	,000
	N		40	40	40	40

** . La corrélation est significative au niveau 0.01 (bilatéral).

Annexe 6: temps moyen par page. Etude du coefficient de corrélation

Corrélations

			nb_pg_vu	retour_s	achat	recomman
nb_pg_vu	Corrélation de		1	,124	,121	,165
	Pearson					
	Sig. (bilatérale)			,373	,383	,232
	N		54	54	54	54
retour_s	Corrélation de		,124	1	,553**	,694**
	Pearson					
	Sig. (bilatérale)		,373		,000	,000
	N		54	54	54	54
achat	Corrélation de		,121	,553**	1	,442**
	Pearson					
	Sig. (bilatérale)		,383	,000		,001
	N		54	54	54	54
recomman	Corrélation de		,165	,694**	,442**	1
	Pearson					
	Sig. (bilatérale)		,232	,000	,001	
	N		54	54	54	54

** . La corrélation est significative au niveau 0.01 (bilatéral).

Annexe 7: Nombre de pages vues. Etude du coefficient de corrélation

