



Modélisation des parcours de navigation

Fabrice Le Guel

CREM, Université de Rennes 1
ADIS, Université Paris-Sud 11

Thierry Pénard

CREM, Université de Rennes 1

1 juillet 2005

<http://www.marsouin.org>



Analyser les données de navigation des internautes, afin :

- (3) d'identifier des parcours/comportements types ou génériques,**
- (2) de comprendre les déterminants de ces parcours et de mieux prédire leurs évolutions**

Intérêt technique et commercial pour les acteurs économiques de l'Internet

Intérêt pour l'action publique (autour de la fracture numérique)

■ La littérature dans ce domaine

Une littérature pluridisciplinaire qui analyse les comportements de navigation

- En **informatique** (Glassman, 1994; Catledge & Pitkow, 1995...) [gestion du réseau Internet]
- En **sociologie** (Hargittai, 2002, 2004, 2005; Lebart, Licoppe, Beaudouin, Wang, Hawk & Tenopir, 2000) [fracture numérique de 'premier' et 'second' niveau]
- En **économie et marketing** [modélisation des choix en ligne]

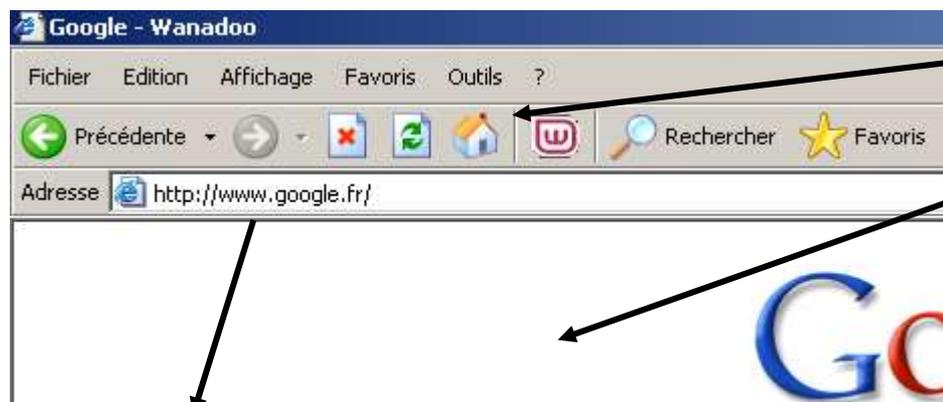
■ La littérature en économie et les programmes de recherche

L'utilisation des données de navigation dans l'analyse économique et marketing :

Usages (choix) / Niveau d'utilisation (intra-inter)	Comportements intra-site	Comportements inter-sites
Recherche d'information/navigation	<p>Choix entre différentes pages Bucklin & Sismeiro (2003) Huberman & al. (1998) Lee, Zufryden, Drèze (2001) ...</p>	<p>Choix entre différents sites Goldfarb (2002a, 2002b, 2003, 2004, 2005) Lee, Drèze, Zufryden (2003a, 2003b) Montgomery & Faloutsos (2000) Montgomery (1999) ...</p>
Achats en ligne	<p>Choix entre différents produits (commerce électronique) Johnson, Bellman, Lohse (2003) Moe & Fader (2004a, 2004b, 2005) Johnson et al. (2004) Lynch & Ariely (2000) Mandel & Johnson (2002) Fader & Hardie (2001) Montgomery & al. (2004b) Moe (2003) ...</p>	<p>Choix entre des produits homogènes (Shopbots) Smith & Brynjolfsson (2001) Brynjolfsson & Smith (2001) Montgomery & al. (2004, 2005) ...</p>

■ Méthodologie pour l'étude des comportements de navigation

1. Utilisation des données de navigation (clickstream data)



Navigateur Internet

Page Internet

Les données de navigation correspondent donc aux enregistrements de l'ensemble des 'clics' (pages vues/visites de sites) de l'internaute sur une session donnée

Adresse URL (Uniform Resource Locator)

Principe : chaque 'clic' (chaque page vue) déclare l'enregistrement de l'adresse URL dans un fichier

```
csl7 787125748 56906 "http://www.ludvigsen.dhalden.no/webdoc/ahlan.gif" 1700 2.691705
csl7 787125764 828867 "http://www.ludvigsen.dhalden.no/webdoc/ahlan.gif" 0 0.0
csl7 787125785 497644 "ftp://ftp.hiof.no/pub/levant/lebanon/jbeil/jbeil01.jpg" 21839 15.229143
csl7 787126872 204994 "http://www.ludvigsen.dhalden.no/webdoc/levant_servers.htm#Lebanon" 0 0.0
csl7 787126872 389659 "http://www.ludvigsen.dhalden.no/webdoc/ahlan.gif" 0 0.0
csl7 787128239 578430 "ftp://ftp.hiof.no/pub/levant/music_and_song/music.list" 40418 8.752783
csl7 787128301 795825 "http://www.ludvigsen.dhalden.no/webdoc/levant_servers.htm#Lebanon" 0 0.0
csl7 787128302 100752 "http://www.ludvigsen.dhalden.no/webdoc/ahlan.gif" 0 0.0
csl7 787128330 9639 "ftp://ftp.hiof.no/pub/levant/music_and_song/singers/AbdelHalimHafez/abdh.txt" 1385 3.227795
csl7 787128350 941071 "http://www.ludvigsen.dhalden.no/webdoc/levant_servers.htm#Lebanon" 0 0.0
csl7 787128351 203370 "http://www.ludvigsen.dhalden.no/webdoc/ahlan.gif" 0 0.0
csl7 787132565 117453 "http://www.cpsr.org/dox/home.html" 3232 3.803559
csl7 787132569 78375 "http://www.cpsr.org/dox/cpsr.gif" 1644 8.892870
csl7 787132588 528202 "http://www.cpsr.org/dox/memb.form.html" 2644 3.432074
```

■ Méthodologie pour l'étude des comportements de navigation

2. Développement d'un logiciel de recueil des données de navigation : **eXante** (Exécutable pour l'Analyse des TracEs)



3. Définition d'une problématique précise liée aux comportements de navigation :

La 'fracture des usages' sur Internet
(capacité des internautes à remplir une

Questions : tâche particulière)

- Existe-t-il autant de fractures que d'usages en ligne ?
- Quelles peuvent être les origines des ces fractures ?

Implications :

- Réduction de la fracture numérique de 'second niveau' (celle des usages en ligne)
- Compréhension de la dynamique d'adoption et de diffusion d'un site Internet

■ Méthodologie pour l'étude des comportements de navigation

4. Pour répondre à ces questions :
mise en place d'une
expérience dans un
laboratoire informatique (le
LABEX, Faculté des
Sciences Économiques,
Université de Rennes 1,
octobre/novembre 2004)

On demande à 50 étudiants
volontaires:

2. D'effectuer 8 tâches en ligne (en observant leurs comportements de navigation grâce à la sonde eXante)
3. De répondre à deux questionnaires (profils socio-économiques, expérience relativement à Internet...).

Tâche 3 [INDICE des PRIX]: trouvez l'indice des prix à la consommation (hors tabac) pour l'ensemble des ménages de juillet 2004 (base 100 en 1998).

Tâche 4 [ACTION FT]: trouvez le cours de l'action de France Télécom à la Bourse de Paris : 1. à l'ouverture aujourd'hui 2. à l'ouverture lundi dernier (le 8/11/2004).

Tâche 5 : [METEO] trouvez un site permettant de consulter les prévisions météo et donnez le temps prévu pour mardi prochain à Quimper.

Tâche 6 : [CAPES] quel est le nombre de postes au concours du CAPES externe de Sciences Economiques et Sociales pour l'année 2004 et 2005.

Tâche 7 : [eADMINISTRATION] faites une demande de certificat de situation du véhicule appartenant à Fabrice Le Guel, immatriculé 1368 WS 44, date de première mise en circulation le 25 juillet 1989.

Tâche 8 : [CINEMA] vous voulez aller au cinéma ce soir à Nantes. Trouvez la salle et les heures de séance pour le film suivant : 'Un long dimanche de fiançailles' (à partir de 19 heures, un cinéma suffit).

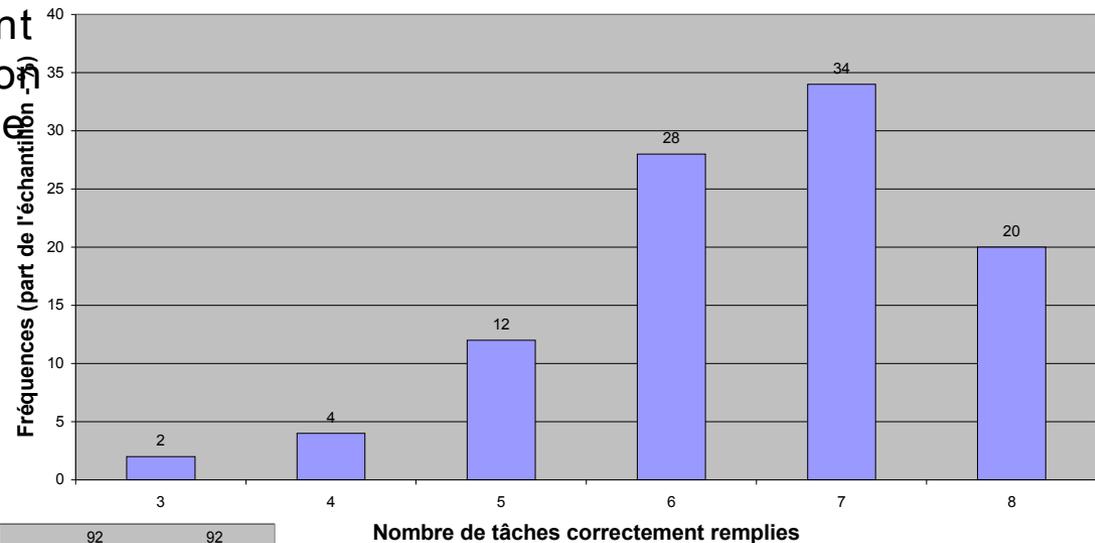
Tâche 9 : [eCOMMERCE1] trouvez deux offres tarifaires pour l'appareil photo numérique neuf Canon PowerShot A85 à moins de 270 euros frais de port inclus (hors accessoires et hors garanties supplémentaires). Précisez le nom du site ou du cybercommerçant.

Tâche 10 : [eCOMMERCE2] trouvez deux offres tarifaires pour le film Podium DVD (édition simple) à moins de 21 euros frais de port inclus. Précisez le nom du site ou du cybercommerçant.

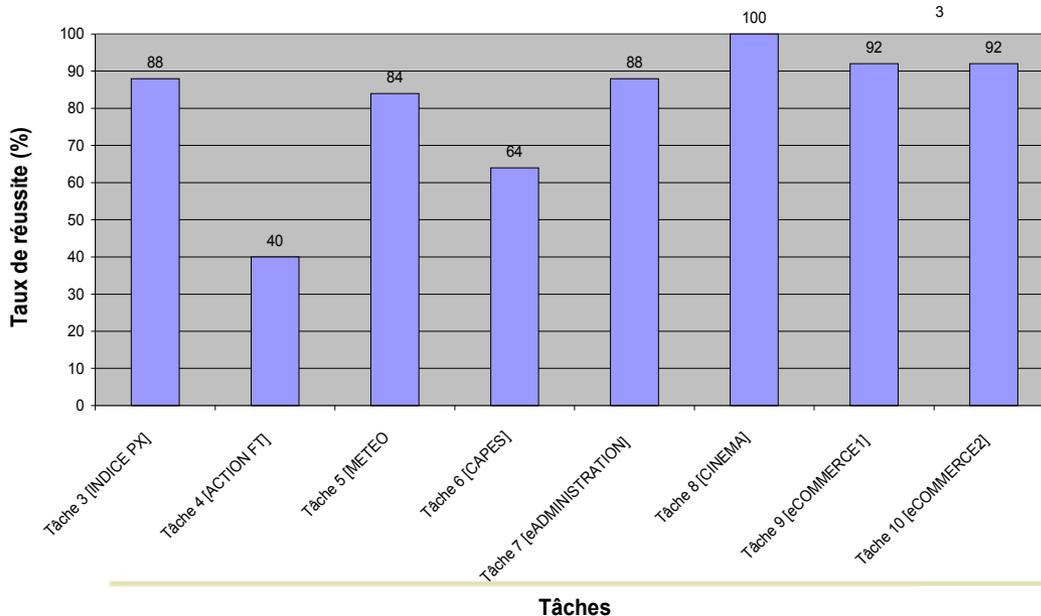
■ Résultats (1) : une fracture des usages...

Les données de navigation + les réponses aux questionnaires sont deux types d'informations que l'on croise pour effectuer une analyse économique

Distribution (à travers l'échantillon) du nombre de tâches correctement remplies



Taux de réussite pour chaque tâche



Au moins deux groupes d'utilisateurs se détachent : ceux qui réussissent pratiquement toutes les tâches, ceux qui réussissent peu de tâches

■ Résultats (2) : ...une fracture des usages faiblement expliquée par les variables socioéconomiques traditionnelles

Les variables qui comptent :

- l'entourage des internautes (capacité à recevoir l'aide d'un tiers)
- l'expérience de navigation pour la tâche demandée (capacité d'apprentissage individuelle, capacité à utiliser correctement un moteur de recherche)

Les raisons des fractures ?

- le manque d'expérience/de formation des internautes
- le manque d'aide d'un tiers
- une mauvaise ergonomie/organisation du site Internet

Conséquences :

- les actions visant à former et aider les internautes débutants comptent
- sans cela, seuls les internautes à forte capacité d'apprentissage sont capables de 'consommer' une grande partie de l'Internet (à l'inverse, situation de blocage pour les internautes les plus faibles)

Modélisation des comportements de navigation

- Des premiers résultats obtenus par des méthodes factorielles et de classification (F. Brucker, Lussi)
 - Existence d'un parcours quasi-identique d'un côté et de parcours très hétérogènes de l'autre
 - Google au cœur des parcours (84% de PDM et de 11 à 30% des sites visités)
- Développement de la sonde eXante (application industrielle)
 - Analyses des usages pour l'ensemble des biens numériques (téléphones portables, Lecteur MP3, TNT...)
 - Initier d'autres expériences sur les usages en ligne



Web Use Project

[Eszter Hargitta](#)'s research lab at Northwestern University is called the Web Use Project. The goal of the [research](#) conducted in this lab is to learn about how people use the Web in their everyday lives and in particular, how differences in Internet use may contribute to social inequality.

