

# Parcours et territoires personnels sur le Web

Séminaire « Flux Internet II »  
23/09/2004

Thomas Beauvisage

## → Projet SensNet

⑦ partenariat FTR&D, Nielsen-NetRatings, Paris III, CNRS-LIMSI

⑦ objectifs généraux :

- analyse de l'offre de contenu sur le Web : décrire et classer les pages et les sites Web
- analyse des usages de l'internet à domicile

## → Thèse (soutenance oct. 2004)

## → données de trafic centrées utilisateur

- ⑦ sonde de recueil de données installée sur le poste des utilisateur (sonde NetRatings)
- ⑦ Couverture de tous les protocoles : Mail (SMTP/POP3), Web (HTTP), News (NNTP), échange de fichiers (FTP) + autres protocoles audio/vidéo...
- ⑦ données Web : liste horodatée de toutes les adresses des pages vues pour chaque utilisateur

## → panel longitudinal : **606 individus** suivis **trois ans**, de janvier 2000 à octobre 2002

- durée d'observation inédite
- observer l'évolution des pratiques, la construction des territoires, la stabilisation des usages

	SN00-02
Nb utilisateurs	606
Nb sessions	261 634
Durée d'observation	34 mois
Nb pages vues (en millions)	15,66
Nb pages différents (en millions)	4,6

## → Les sites vus par l'utilisateur :

- ⑦ la fréquence, la régularité de visite
- ⑦ quels contenus pour quelle fréquence
- ⑦ quelle « morphologie » de navigation en fonction de ces territoires

## → Vue longitudinale

- ⑦ comment évolue le « corpus » de sites des internautes
- ⑦ quelles sont les modalités de découverte, de balisage, etc
- ⑦ comment sont valorisées les différentes « zones » du Web et comment elle s'articulent au sein de la navigation

## → esquisse une éthologie du Web tel qu'il peut être parcouru et s'insère dans les pratiques

# Audience globale



→ Globalement, peu de sites vus par plusieurs panélistes

- ⑦ 2/3 des 192 000 sites ne sont vus que par une personne sur les 3 ans
- ⑦ une trentaine seulement sont visités par plus de la moitié du panel
- ⑦ on retrouve dans cette liste les grands portails généralistes, les sites de FAI et les sites de services
- ⑦ reflète des centres d'intérêt variés (cf expé BibUsages : 76 individus sur 6 mois, beaucoup plus de sites en commun)

## Nombre de panélistes différents par site/portail

	Nombre de sites	Part des sites
1 visiteur	127 357	66,3 %
2 à 5	49 446	25,8 %
6 à 99	14 818	7,7 %
Plus de 100	378	0,2 %
<i>Total</i>	<i>192 000</i>	<i>100%</i>

# Territoires personnels



## → A l'échelle de l'utilisateur

- ⑦ 943 sites différents par utilisateur en moy. sur les 34 mois d'observation (médiane : 569 - min/max : 11 / 8900)
- ⑦ en moyenne, les  $\frac{3}{4}$  des sites d'un internaute n'apparaissent que dans une seule session
- ⑦ cette structure est constante quelle que soit l'intensité d'usage (les utilisateurs intensifs ne « surfent » comparativement pas plus que les autres)
- ⑦ seule une poignée de sites figurent dans plus de 40 % des sessions de chaque internaute

### Intensité d'usage et présence des sites visités dans les sessions des internautes

Quartile du nombre de sessions total	Sites vus dans 1 session du pan.	Sites vus dans plusieurs sessions du pan.			Total
		0 à 19 % de ses sessions	20 à 39 % de ses sessions	40 à 100 % de ses sessions	
1er quartile	78,0 %	18,8 %	1,7 %	1,5 %	100%
2e quartile	75,7 %	23,3 %	0,5 %	0,5 %	100%
3e quartile	73,7 %	25,8 %	0,3 %	0,2 %	100%
4e quartile	71,3 %	28,4 %	0,8 %	0,1 %	100%

( diffusion libre )

# Vue longitudinale



## → corpus en constante expansion

- ⑦ les panélistes ne cessent de voir des nouveaux sites tout au long de la période d'observation : quasi-linéaire entre sessions et corpus visité
- ⑦ les « sites-hapax » ne sont pas liés à une période de découverte préliminaire

## → empan des sites

- ⑦ fréquence vs. régularité → l'empan comme signe de fidélité et d'ancrage dans les usages  
Cf : certains sites vus systématiquement dans des contextes rares (préparation des vacances, télédéclaration d'impôts, etc.)
- ⑦ empan des sites vus dans plus d'une session :
  - 1/3 des sites < 30 jours
  - 1/4 des sites > 1 an
- ⑦ structure constante quel que soit le nombre total de sites visités : 8 % de l'ensemble des sites vus ont un empan de plus d'un an

# Structuration du territoire



## → Trois zones-type

- ⑦ sites « éphémères », vus dans une seule session : 75% du corpus
- ⑦ sites vus plus d'une fois, mais qui ne fidélisent pas l'internaute : 17% du corpus. Ils sont actifs sur des périodes relativement courtes, de l'ordre d'un mois, et correspondent à des activités bornées dans le temps (ex: recherche d'appartement, recherche sur un thème occasionnel...)
- ⑦ sites familiers, d'empan supérieur à un an : 8% du corpus. On y retrouve les portails généralistes et les moteurs, ainsi que les services qui fidélisent les internautes (WebMail, notamment) + des sites relatifs à des thèmes propres à l'utilisateur.

## → Structure globalement partagée par l'ensemble des panélistes

# Du familier vers l'inconnu



➔ les sites habituels sont très minoritaires, mais ils structurent l'ensemble de la navigation

⑦ les sites dont l'empan dépasse un an sont présents dans 91 % des sessions

## % de sessions contenant au moins un site d'empan..

Empan nul (site vu dans une seule journée)	43%
1 à 9 jours	13%
10 à 30 jours	12%
Un à cinq mois	35%
Six à 12 mois	36%
Plus d'un an	91%

⑦ dans plus des 2/3 des sessions, c'est sur ces sites que l'internaute passe le plus de temps dans la session

Catégorie de sites la plus présente dans la session en durée	% sessions
Sites vus dans une session (empan nul)	9,5 %
Sites occasionnels, très faible durée de vie (empan de 1 à 9 jours)	2,0 %
Sites occasionnels, faible durée de vie (empan de 10 à 29 jours)	1,6 %
Sites occasionnels, durée de vie moyenne (empan de 1 à 5 mois)	8,2 %
Sites occasionnels, durée de vie importante (empan de 6 à 12 mois)	8,7 %
Sites habituels (empan de plus d'un an)	70,1 %



# Territoires thématiques



- caractérisation du contenu des pages vues par
- ⑦ les annuaires du Web : utilisation de la position du site vu dans la catégorie de premier niveau
  - ⑦ CatService (FTRD), outil de catégorisation semi-automatique de pages, ici appliqué aux portails généralistes et aux services Web (comm., recherche)
  - ⑦ catégorie « pornographique » dans les données de base
- 30 descripteurs (pour Nomade)

## Catégories CatService

Services Web
Moteur
WebMail
WebChat
Portails généraliste : services
Achat
Aide
Annuaire
Bourse
Communication
Divers
Généralités
Information Produit
Information Service
Informations
Loisir En Ligne
Non catégorisé
Page Accueil
Page Perso
Personnalisation

## Catégories d'annuaires

Nomade (12 catégories)
Actu, médias
Culture et loisirs
Éducation, formation
Espace B to B
Forme et Santé
Mes Courses
Nature et sciences
Nouvelles technologies
Société, Vie pratique
Sorties, spectacles
Sport et détente
Voyage, géographie

# Territoires thématiques



## → Cartographie des thèmes et services accédés par chaque internaute

- ⑦ globalement, les sessions sont mono-thème / mono-service : une catégorie décrit la majorité de la durée de la session
- ⑦ forte concentration thématique des visites au niveau individuel : en moyenne, 5 catégories décrivent 95 % du contenu des sessions d'un internaute.
  - Sur l'ensemble du trafic de chaque panéliste, un faible nombre de catégories décrit un grand nombre de sessions.
- ⑦ les utilisateurs les plus intensifs sont les plus diversifiés

# Classification



## → Typologie des sessions en croisant

- ⑦ la fréquence des sites dans les sessions de l'utilisateur
- ⑦ l'empan des sites de la session
- ⑦ l'importance des thèmes / services accédés

Famille de variable	% de la durée de la session sur des...
Fréquence des sites	sites n'apparaissant que dans cette session sites apparaissant dans 0 à 9 % des sessions sites apparaissant dans 10 à 39 % des sessions sites apparaissant dans 40 à 100 % des sessions
Empan des sites	sites n'apparaissant que dans une journée (empan nul) sites d'empan entre 1 et 29 jours sites d'empan entre 1 et 6 mois sites d'empan entre 7 et 12 mois sites d'empan supérieur à un an
Fréquence des thèmes / services	thèmes / services très rares thèmes / services rares thèmes / services occasionnels thèmes / services fréquents thèmes / services habituels

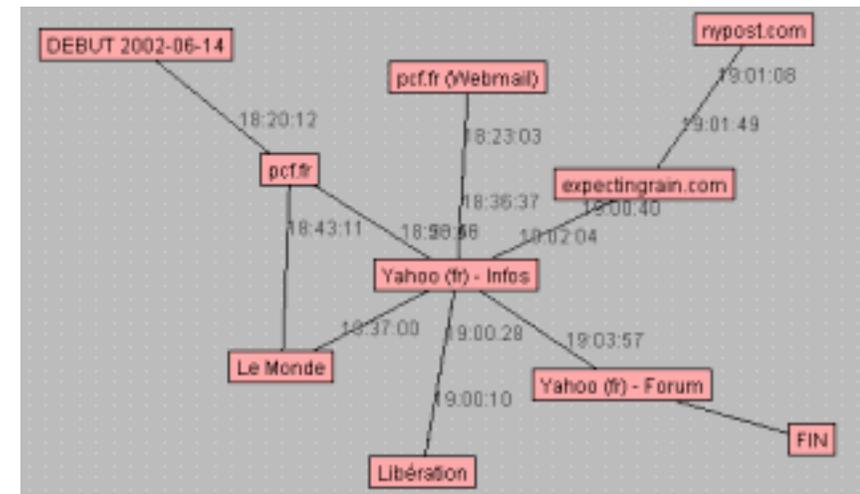
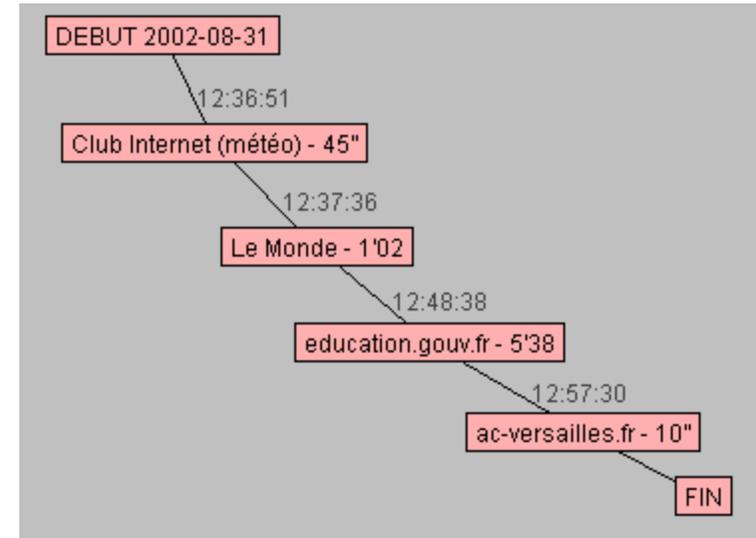
## → 3 classes apparaissent

- ⑦ parcours routiniers (52 %)
- ⑦ parcours « contenus occasionnels » (36 %)
- ⑦ parcours exploratoires (12 %)

# Sessions routinières



- ➔ contenus très fréquents, et de fort empan
  - ⑦ 93 % de la durée des sessions sur des sites d'empan > 1 an
  - ⑦ 58 % sur des sites figurant parmi les plus fréquents
  - ⑦ thèmes et services fréquents/habituels sur-représentés
  
- ➔ parcours globalement plus courts et plus linéaires que les autres, articulant plus ou moins des contenus fréquemment visités (cf exemple)
  
- ➔ sur le plan des contenus, ces parcours sont orientés vers des pages renouvelées fréquemment (sites d'information, services de communication, bourse en ligne, etc.)
  - ➔ flux d'information et de communication



# Parcours « contenus occasionnels »

- sites d'empan plus faible (1 mois à 1 an), apparition de thèmes peu fréquents
- parcours plus complexe : détours, feuilletage à l'intérieur des sites
- contenus plus ouverts
  - ⑦ plus rares pour l'utilisateur
  - ⑦ usage des moteurs, mais pour des requêtes ponctuelles et ciblées (ex : requête « allocine » sur Yahoo), pseudo-bookmark
  - ⑦ orientation vers des contenus de type services, pages jaunes, galerie marchande : sites de voyage, courses, ...

Catégorie dominante dans la session	Sessions « contenus occasionnels »	Ensemble des sessions
Mes Courses	<b>12,5 %</b>	7,2 %
Société, Vie pratique	<b>11,6 %</b>	9,6 %
PORNO	<b>10,2 %</b>	5,9 %
Sport et détente	<b>9,6 %</b>	7,3 %
Espace B to B	8,9 %	11,3 %
Actu, médias	8,3 %	5,0 %
CS – WebMail	6,1 %	12,7 %
Nouvelles technologies	5,0 %	2,8 %
CS – Moteur	4,0 %	3,7 %
Culture et loisirs	3,7 %	3,0 %
PG - Page Accueil	3,4 %	12,5 %
Voyage, géographie	3,4 %	2,1 %
Autres	13,5%	16,9 %

# Parcours exploratoires



→ coloration de découverte

⑦ sur-représentation des sites-hapax et des catégories thématiques rares

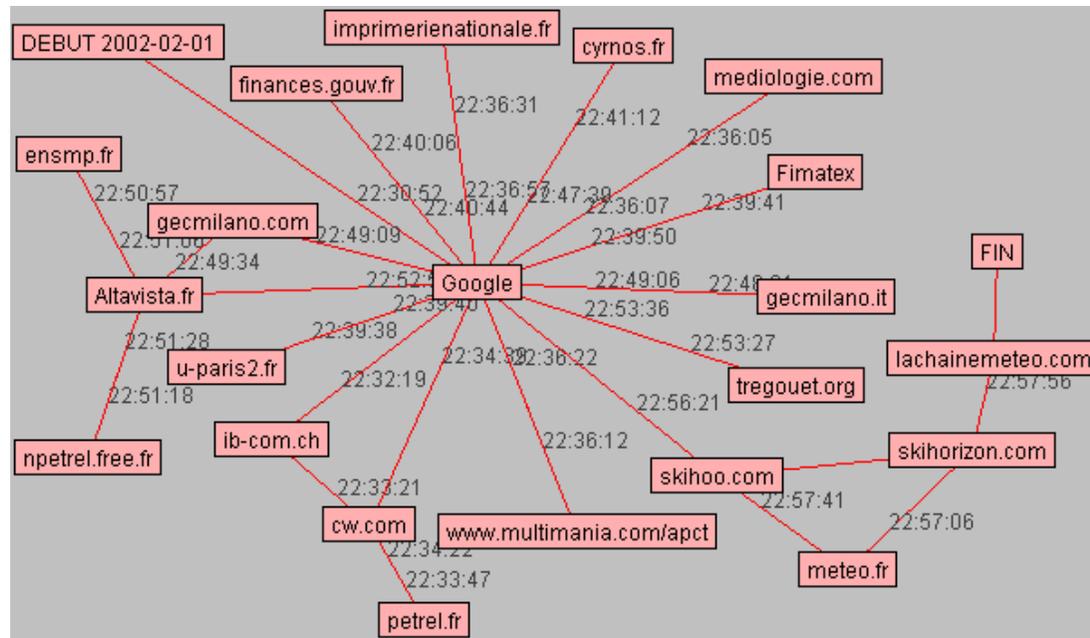
→ effet de recherche et de feuilletage

⑦ formes très éclatées de parcours, avec beaucoup de pages-pivots

⑦ fort usage des moteurs

⑦ peu de temps sur chaque site

→ Contenus : Web comme source d'informations, orientation vers les documents



( diffusion libre )

# Conclusion



- Structuration du territoire comparable chez l'ensemble des panéliste, avec des variations d'intensité
  
- Diversité entre les utilisateurs
  - ⑦ disparité des centres d'intérêt
  - ⑦ la valorisation des contenus dépend de l'utilisateur
  
- Toutefois, en termes fonctionnels, les types de contenus sont liés aux types de parcours
  
- prendre en compte la spécificité des contenus du Web, et ne pas les voir uniquement en termes de documents, mais également en termes de flux, de services et d'outillage.