

Réseau Thématique Pluridisciplinaire n° 35
Economie, Organisation et TIC

Action Spécifique :
Les mesures de flux Internet

Responsables :

*Valérie Beaudouin**, *Ludovic Lebart***

* France Télécom- Recherche Développement,

** CNRS, ENST

Les mesures de flux Internet

- 0. Cadre spécifique
- 1. Captation des flux
- 2. Analyse des traces
- 3. Mesures de flux sur individus
- 4. Mesures sur site
- 5. Mesure du réseau lui-même

Cadre Spécifique

- Evolution rapide, Manque de recul \Leftrightarrow Veille technologique
- Mesures « quasi-gratuites » et pléthoriques
- Importantes lacunes
- Variété des enjeux
- Contexte de bouleversement des pratiques

1. La captation des flux

- « Traces d'usage »
- Sondes sur le poste de l'utilisateur (approche *user centric*)
- Les systèmes de recueils de logs sur les serveurs de sites (approche *site centric*)
- Sondes sur réseaux (mesure des flux, volumes)

2. L'analyse des traces

- Pré-traitement des traces brutes : filtrages, construction d'indicateurs (pages vues, sessions, durée...), catégorisation des services, analyse des contenus textuels...
- Les **hits** (chargement d'un fichier) moins utilisés que les **pages** ou les **visites** (ensembles de pages lues sur un même site : < 30 minutes entre 2 pages).
- Importance du **temps passé** (arbitrage entre activités sous contrainte, styles de vie).
- Mesures en **bits** ou bytes (ordres de grandeurs complémentaires des précédents). Elles servent à dimensionner les capacités des réseaux.

2. L'analyse des traces (*suite*)

- Paramètres dérivés (nombres de pages / visite ou / heure, nombre de secondes par page, de sites par session ...)
- Mesures plus complexes (trajectoires, profils de sites par session).
- Bases de données, d'images et de sons constituées par les séquences (ordonnées ou non) des pages lues. (partie texte des pages HTML / XML)
- Relevés donnant lieu à des traitements exploratoires (Data Mining, Text Mining).
- Qualification des visiteurs d'une machine connectée à partir d'analyses textuelles ou sémantiques des pages lues.

3. Mesures de flux sur individus (*user centric*)

- Mesures d'audience, cf. Instituts de sondage (ex : panels d'internaute NetValue, Nielsen-NetRatings, Jupiter MMXI, et actuellement NetRatings).
- Enquêtes de cadrage \Rightarrow redressements.
- Recueil passif à l'aide de logiciels installés.
- Instance interprofessionnelle de contrôle : **CESP** (Centre d'Etude des Supports de Publicité : audite les audiences radio, télévision, presse, cinéma, affichage).
- L'enjeu : tarifs publicitaires \Rightarrow valorisation boursière.
- Le média internet \cong 1% des dépenses de publicité.
- Difficulté des mesures d'audience en entreprise.
- Impliquer les instituts dans des programmes coopératifs.

4. Les mesures sur site (*site centric*)

- → à partir des **fichiers logs** (auxquels échappent les échanges via **proxys**)
- → à partir de **marqueurs (tags)** attachés à chaque page: Comptabilisation des chargements quelle que soit leur origine.
- Instance interprofessionnelle de contrôle : **BIM** (Bureau Internet Multimedia), sous l'égide de « Diffusion Contrôle ».
- D'autres mesures, intra-sites cette fois, sont destinées à l'optimisation de la conception et de la gestion de ces sites (cf. Action **AXIS** de l'INRIA : *Conception, analyse et amélioration de systèmes d'information dirigées par l'usage*).
- Elles nécessitent la conception de logiciels spécifiques.

5. Mesure du réseau lui-même

On distingue :

- les **mesures statiques** (architecture du réseau), le réseau pouvant être décrit par des logiciels d'exploration (**crawler**).
- les **mesures dynamiques** (prenant en compte les navigations et les parcours des utilisateurs), incorporent les mesures liées à l'interactivité. (ex : échanges pair à pair (**peer to peer**) pour des produits culturels).
- exploitation des possibilités de **marketing direct** de la toile (exploitation **en temps réel** de toutes les techniques de ciblage, géo-marketing, vente par correspondance)
- profils d'achat enrichis par les **profils sémantique** de lecture des pages, les parcours, et les profils sémantiques des produits commandés.

5. Mesure du réseau lui-même (*suite 1*)

- Les niveaux et types de mesures les plus novatrices concernent
- les mesures locales (locales sur le graphe, et non géographiquement) (communautés, monographies)
- Les communautés d'échange, d'expérience, de création (ou épistémique) (cf. travaux sur l'économie des communautés médiatés).
- Ces mesures ne se limitent pas à la toile, mais concerne l'ensemble des protocoles et messageries.

5. Mesure du réseau lui-même (*suite 2*)

- Plus ambitieuse encore est l'étude des possibles réorganisations ou structurations de l'ensemble du tissu social : relations personnelles, relations professionnelles, déplacements, groupes sociaux flous ou empiétants (graphes non limités à des ensembles de cliques).
- L'extension attendue de la toile laisse entrevoir de profondes transformations des catégories d'analyse et des outils qui leurs sont associés.

Conclusion (*à ce stade ...*)

- Chaque thème cité doit être compris comme un sujet de recherche potentiel.
- Il reste à articuler ces travaux sur les mesures de flux avec les thèmes d'autres actions spécifiques, qu'il s'agisse de la **définition de nomenclatures concernant les nouveaux biens numériques**, d'étude des **effets des flux d'échanges internationaux** de ces biens numériques, de **modèles plus expérimentaux d'interactivité**.