

Métrologie des interactions et Services adaptables

Luigi Lancieri

<http://www.lifl.fr/~lancieri/>

Luigi.lancieri@univ-lille1.fr

Plan

1. *Introduction*
2. *Modélisation*
3. **Métrologie des interactions et Services adaptables**
 - Instrumentation, collecte et analyse de traces
 - Indicateurs de comportements
 - Systèmes d'alertes et de recommandations

Collecte et analyse de traces

Traces et Interactions

- Sur le plan technologique, une médiation plus ou moins transparente.
- Granularité variables.
 - Fichiers de log, cookies, Artefacts, contenus consultés (textes, images,....)
- Plus ou moins accessibles
 - Localisation, autorisations

Différentes approches

- Approche écologique
 - Exploiter les traces existantes (serveurs web, proxy, données d'accounting, ..)
 - **Faire avec ce que l'on a.**
- Approches dédiées
 - Identifier les instruments, les lieux, les cibles de la capture (camera, capteurs biométriques, déplacement, réseaux de capteurs ..)
 - Bien cerner ce que l'on fera avec les données (coûts liés à la volumétrie, traitement, aspects éthiques et juridiques, ..)

Localisation des « capteurs »

- Obtenir des résultats de type quantitatif ou sémantique
 - **Quantitatif**: nombre de requêtes par seconde sur un serveur
 - **Sémantique**: l'utilisateur Dupond est surtout intéressé par la pêche à la ligne.

•Coté Client

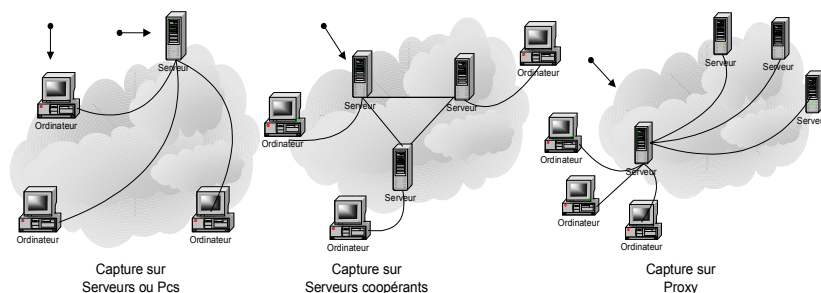
- Sondes, spyware,...
- Lourd à mettre en œuvre
- Ne représente qu'une personne
- Frein à la spontanéité

•Coté Serveur

- Simple à mettre en œuvre
- Bonne représentativité de l'intérêt du contenu du serveur
- La représentativité des statistiques sur le comportement des usagers est limitée

•Coté médiateur (e.g proxy cache)

- Simple à mettre en œuvre
- Bonne représentativité des statistiques utilisateurs
- Représentativité limitée des statistiques sur le contenu d'un serveur donné.



Quelques exemples concrets

- Web services et API
 - Facebook, tweeter, google ..
- Interactions en ligne
 - Fichiers de log : Pages visitées, mails lus (images)
 - Cookies,
- Interactions dans un environnements physique
 - Données backoffice, carte de fidélité, achats
- Capteurs en magasin
 - Temps passé dans un rayon, parcours

Étude de cas : la mesure du déplacement

- Avec référence
 - Analyse par corrélation d'images et suivi d'objets (eg caméra dans les magasins)
 - Handover bornes radio (wifi, umts, ..)
 - Retour d'information gps
 - Capteurs optiques de proximité, sens de déplacement
- Sans références
 - Accéléromètres sur smartphone
 -

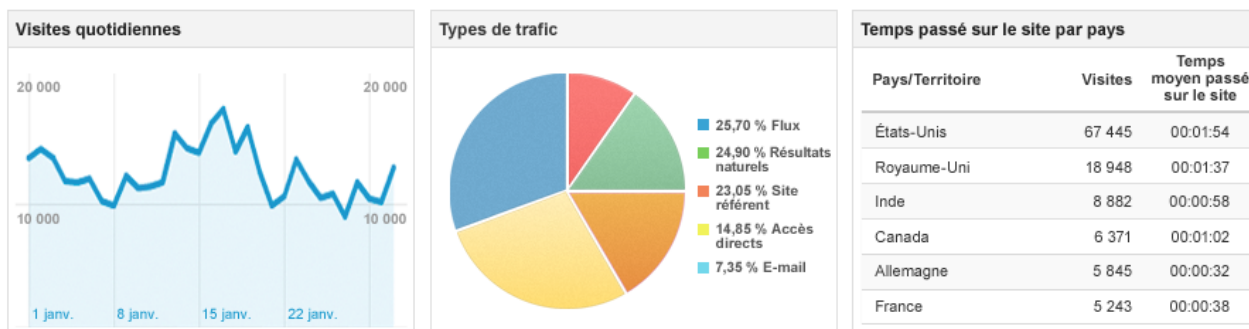
Étude de cas

Google Analytics

- Permet de suivre l'activité de votre site
 - Version évolué des compteurs d'accès
 - Le potentiel peut être très important surtout si associé à adwords
- Facile à mettre en œuvre, utile pour des petites entreprises.
- Google profite aussi de l'information et peut être intéressé comme médiateur publicitaire

Google Analytics (2)

Mon tableau de bord



Activités



Google Analytics (2)

- Après l'inscription gratuite, Google fournit un script à copier entre les balise <HEAD> de chacune des pages à surveiller. Plusieurs plugins disponibles (Joomla, Wordpress,..)
- Les statistiques sont disponibles au plus 24h après l'installation, puis ensuite disponible en quasi temps réel.
- Vidéo de présentation par Google (9 min)

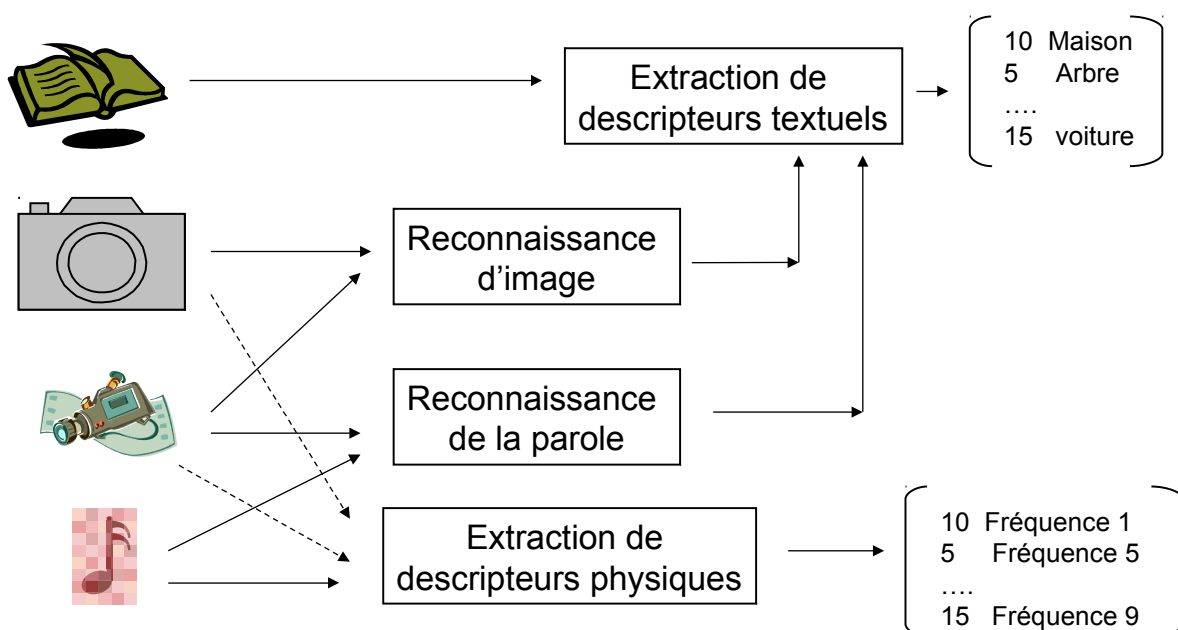


Des traces aux indicateurs

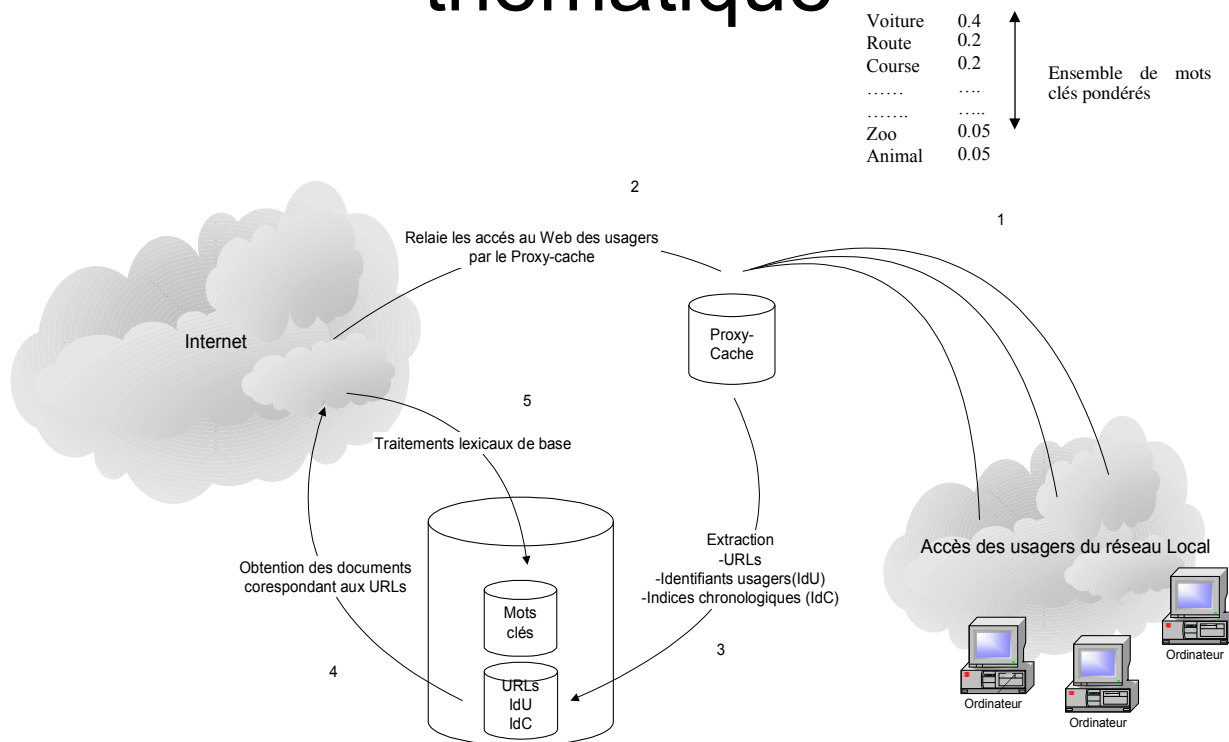
- Identifier des sources de traces utiles est un problème en soi.
- Utiliser ces données pour construire des indicateurs de comportements **réalistes** est tout aussi problématique.
- Valider les indicateurs, par exemple, en vérifiant leurs caractères prédictifs.

Exemple de profil d'usager

- Par la caractérisation d'objets multimédias



Exemple d'obtention d'un profil thématique



Aspects éthiques et juridiques de l'utilisation de traces

- Historiquement, le problème n'est pas nouveau
 - Pour préserver leur intimité, les premiers usagers du téléphone ne répondaient pas directement lors d'un appel.
 - Les agents des PTT doivent prêter serment
- Principe de base pour l'exploitation de données personnelles: L'utilisateur doit être informé et consentant.
 - La réalité est loin de ce principe qui est difficile à faire respecter sur Internet faute de législation internationale cohérente
- Sources Juridiques
 - Directive Européenne ("Data Protection Directive")
 - Législation Française
 - Norme simplifiée n° 48 : Délibération n° 2012-209 du 21 juin 2012 portant création d'une norme simplifiée concernant les traitements automatisés de données à caractère personnel relatifs à la gestion de clients et de prospects (voir site de la CNIL pour détails)

Les systèmes de recommandations

Introduction

- Une logique duale de celle de la recherche d'information basé sur la prévision
 - La recherche d'information: **je trouve en cherchant**
 - Les services de recommandations: **je trouve sans chercher**
- Une autre approche de l'IHM adaptée quand l'expression de la requête est complexe.
 - Difficile de formuler la requête
 - Problèmes de disponibilités d'IHM suffisamment expressive au bon moment et bon endroit (spontanéité, mobilité, ..)
- Repose sur 2 éléments principaux
 - Donnée caractérisant l'usage (traces, déclarations, ..)
 - Algorithmes (similarités, classification, extractions de règles)

Quelques données

- Amazon déclare qu' 1/3 de ses ventes est généré par son moteur de recommandation.
- Yankee group (2008) indique qu'un moteur de recommandation permet d'augmenter de 5% le revenu des sites en ligne.
- Certaines chaines publicitaires interactives ont vu la consommation de leurs programmes multiplié par 5.
- Video d'application de Kxen chez allociné (6 min)

Types de recommandations

- Éditorial
 - Synthèse, best-of constaté, cible le plus grand nombre.
- Personnalisée
 - Similarité par rapport au profil (déclaré ou par analyse de traces)
- SocialeFiltrage collaboratif
 - Marketing viral, avis réseaux sociaux
- Quelques acteurs (produits, ..)
 - Media unbound (musique), Orca (compass), Epiphany (crm), Kxen
 - **En open source:** Taste , Racofi, lenskit, easyrec

Distances et similarités

- Mesurer la similarité entre deux vecteurs correspondant aux profils ou aux caractéristiques des produits.
 - Distance Euclidienne, Cosinus, Jaccard, etc

Similar due to inheritance

Similar due to... uh... other factors

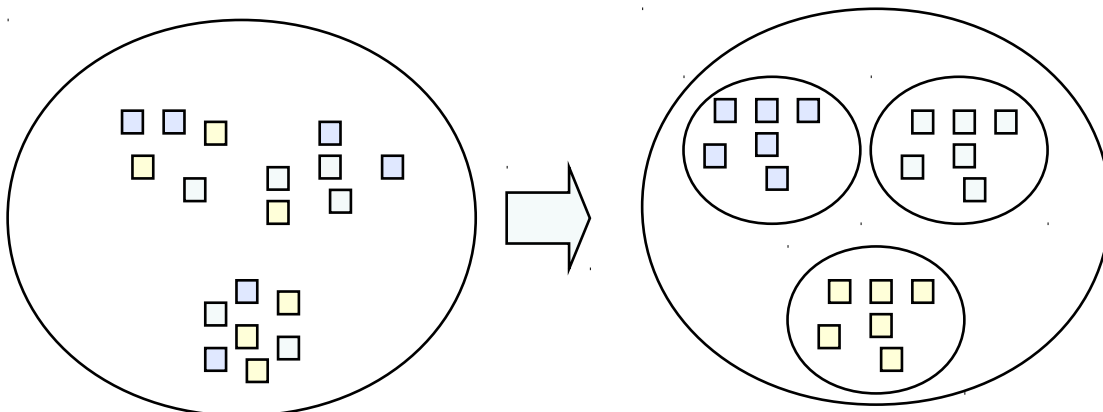


Qu'est ce que la similarité ?

Source evolution.Berkeley.edu

Classification

- Utilisation des distances pour constituer des groupes (eg. communautés d'intérêts)
- Recherche de similarités (ressemblance / dissemblance) entre les items (caractéristiques plus ou moins proches).
- Classification hiérarchique, type K-means, réseaux de neurones, etc



Cas de la distance de Jaccard

$$D = 1 - \frac{(U1 \cap U2)}{(U1 \cup U2)}$$

1- Le rapport entre **taille de l'intersection** sur la **taille de l'union** entre deux ensembles.

4 usagers indiquent
les films qu'ils ont
apprécié



	SpiderMan	StarWars	Colombo	Maigret
User 1	1	1	0	1
User 2	0	1	0	1
User 3	1	1	1	1
User 4	1	0	1	0

U1 est plus proche de U3 que de U2 ou de U4

- $D(U1, U2) = 1 - 2/3 = 0.33$ (starwars, maigret) / (spiderman, starwars, maigret)
- $D(U1, U3) = 1 - 3/4 = 0.25$
- $D(U1, U4) = 1 - 1/4 = 0.75$

Règles d'associations

- Mettre en évidence les items liés
 - Produits achetés ensembles
 - Films correspondant à un même « goût »
- Règles de la forme
 - si produit 1 alors produit 2
 - Si produit 1 et produit 2 alors produit 3
- On mesure la fiabilité (support) et la précision (confiance) des règles

Un exemple simple

Matrice d'utilité

Objectif : Prédire les cases vides de la Matrice d'utilité et surtout les cases avec des **notes** élevées

	SpiderMan	StarWars	Colombo	Maigret
User 1	5	4		1
User 2	???	4		1
User 3	1	2	4	5
User 4	5		1	

User 2 appréciera t'il Spiderman ?

- profils des 2 films proches
- profils de U1 et U2 proches

→ Probablement car il a apprécié SW et les usagers qui ont apprécié SW ont aussi apprécié SM.

Règles de recommandation

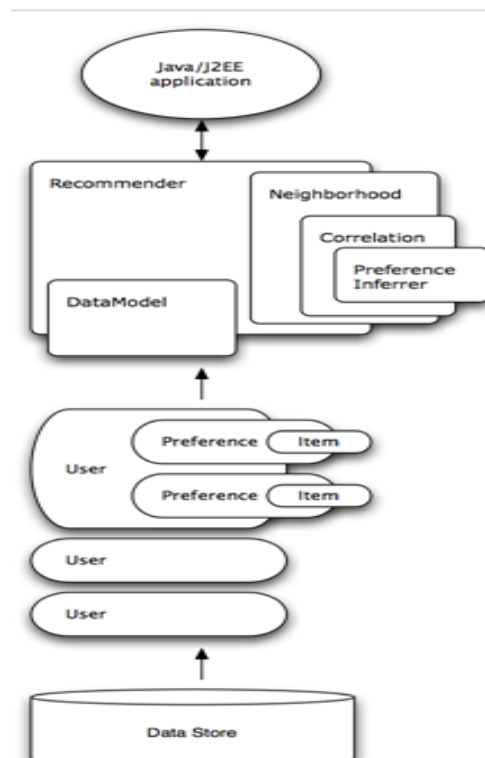
Si SW alors SM
Si C alors M

Matrice de cooccurrence ou de distance

- le nombre de fois où deux éléments sont apparus ensembles avec **note élevée**
- **distance** entre films

	SM	SW	C	M
SM	2	1	0	0
SW		2	0	0
C			1	1
M				1

Exemple d'architecture (Taste)



Liens entre recommandation et modèles théoriques

- Contrairement au commerce physique qui ne vend que des produits populaires (manque de place), le commerce en ligne vend tout, même des produits plus rarement achetés (lois étroite / large).
- Recommandation = réduction orienté sur l'activité de l'espace de donnée (oubli, classification)
- Recommandation et IHM / Media généralisé.
 - Bien adaptée aux IHM sommaires (pas besoin de formaliser, le service prend l'initiative)
 - Bien adaptée aux IHM diffuses ou transparentes