

UE : Personnalisation de l'Information et Analyse de Données Sociales

L'UE est composée de deux (2) matières

1. Personnalisation et Recommandation de l'information
2. Analyse de données sociales

UE : Analyse de Données Sociales et Personnalisation de l'Information



Enseignants

- Personnalisation et Recommandation de l'information

- Systèmes de recherche d'information personnalisée
Lynda Tamine-Lechani (C, TD), lynda.lechani@irit.fr
- Systèmes de recommandation
Gilles Hubert (C, TD, TP), gilles.hubert@irit.fr

- Analyse de données sociales

- Systèmes de gestions de mégadonnées pour les graphes
Gilles Hubert (C, TD, TP), gilles.hubert@irit.fr
- Analyse de réseaux sociaux
Lynda Tamine-Lechani (C, TD, TP) lynda.lechani@irit.fr

Organisation de l'enseignement

11 Séances C/TD (22 H)

- Supports de cours fournis
- TD corrigés en séance
- *TP dans le cadre du projet chef d'oeuvre

4 Séances de TP (8 H)

Modalités de Contrôle des Connaissances (MCC)

1^{ère} session

Contrôle continu (CC)

-CC : Notation d'un compte rendu de TP (*Recommandation de l'information*)

Contrôle terminal (CT)

- Epreuve écrite : 2H

$$\text{Note finale} = 0.3 * \text{CC} + 0.7 * \text{CT}$$

2^{nde} session

Contrôle terminal (CT)

- Epreuve écrite : 2H

- Pas de report de la note de CC de la 1^{ère} session

Personnalisation et Recommandation de l'Information

Bibliographie succincte : Pré-requis en recherche d'information

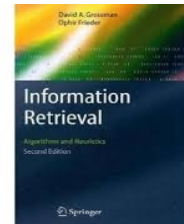
Cours "Recherche d'Information en M1 DC"

Demander le support pour ceux qui ne l'ont pas

Information retrieval: Algorithms and Heuristics

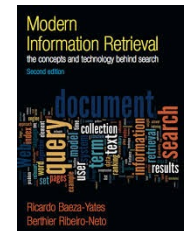
David A. Grossamnn, Ophir Frieder, Kluwer

Academic Publishers, 1998



Modern information retrieval

R.B Yates, R. Neto, ACM Press Addison Wesley, 1999



Recherche d'information, applications, modèles et algorithmes

M.R Amini et E. Gaussier, Eyrolles 2012



Personnalisation et Recommandation de l'Information

Bibliographie succincte : Systèmes d'accès personnalisé à l'information, systèmes de recommandation

Recommender systems:

Charu C. Aggarwal

Springer, 2016



Chapitre : Recherche d'information contextuelle et web

Lynda Tamine, Sylvie Calabretto., p 201-224,

Ouvrage Recherche d'information : Etat des lieux et perspectives

M. Boughanem et Jacques Savoy, Hermes, 1999



Recommender systems, an Introduction

Dietmar Jannach, Markus Zanker,

Alexander Felfernig, Gerhard Friedrich,

Cambridge University press, 2015



Personnalisation et Recommandation de l'Information

Plan du cours

Introduction

- Recommandation et accès personnalisé à l'information : objectifs
- Notion de profil/modèle utilisateur

Partie 1 : Systèmes de recherche d'information personnalisée

1. Modèles de représentation des profils utilisateurs
2. Modèles de recherche d'information personnalisée
3. Evaluation des systèmes de recherche d'information personnalisée

Partie 2 : Protection de la vie privée dans les systèmes de recommandation et systèmes de recherche d'information personnalisée

1. Données sensibles et risques dans les systèmes de recommandation et systèmes de recherche d'information personnalisée
2. Méthodes protection de la vie privée

Partie 3 : Systèmes de recommandation

1. Modèles de recommandation
2. Evaluation des systèmes de recommandation

Objectifs du cours

1. Distinguer les systèmes de recommandation des systèmes de recherche d'information personnalisée
2. Mettre en oeuvre un système de recommandation, système de recherche d'information personnalisée
3. Evaluer les performances d'un système de recherche d'information personnalisée, système de recommandation
4. Proposer et mettre en oeuvre une méthode de protection de vie privée adaptée à une application d'accès personnalisé à l'information (recommandation, recherche d'information personnalisée)

Personnalisation et Recommandation de l'Information

Planning prévisionnel : *Systemes de recherche d'information personnalisée*

Séance	Contenu
S1	Introduction au cours; principes de modélisation de profils
S2	Modélisation des profils (suite)
S3	TD1
S4	Modèles d'accès personnalisé à l'information
S5	Modèles d'accès personnalisé à l'information (suite)
S6	TD2
S7	Evaluation des systèmes d'accès personnalisé à l'information
S8	TD 3
S9	Personnalisation et vie privée

Personnalisation et Recommandation de l'Information

Plan du cours

Introduction

- **Recommandation et accès personnalisé à l'information : objectifs**
- Notion de profil utilisateur

Partie 1 : Systèmes de recherche d'information personnalisée

1. Modèles de représentation des profils utilisateurs
2. Modèles de recherche d'information personnalisée
3. Evaluation des systèmes de recherche d'information personnalisée

Partie 2 : Protection de la vie privée dans les systèmes de recommandation et systèmes de recherche d'information personnalisée

1. Données sensibles et risques dans les systèmes de recommandation et systèmes de recherche d'information personnalisée
2. Méthodes protection de la vie privée

Partie 3 : Systèmes de recommandation

1. Modèles de recommandation
2. Evaluation des systèmes de recommandation

L'information est partout : diversité, volume et multiplicité de sources

De plus en plus d'informations disponibles produites par de plus en plus de sources d'informations : applications sociales, appareils mobiles, capteurs, ...



Quelques chiffres en 2020 :

- En janvier 2020, il y avait plus de **4 milliards d'internautes**.
- L'internaute moyen passe **6 heures et 43 minutes en ligne chaque jour**.
- En janvier 2020, il y avait plus de **1.74 milliards de sites Web sur Internet**

L'information porte sur tout : diversité de contenus et formes

Personne : image et contenu structuré

Barack Obama



Portrait officiel de Barack Obama.

Fonctions

44^e président des États-Unis
20 janvier 2009 – 20 janvier 2017
(8 ans)

Élection 4 novembre 2008
Réélection 6 novembre 2012
Vice-président Joe Biden
Gouvernement Administration Obama
Prédécesseur George W. Bush
Successeur Donald Trump

Nouvelle : image et texte



iPhone 7 128Go
Noir avec un ...
451,99 €
SFR.FR
★★★★★ (1k+)



Jeunesse, parité, société civile: le gouvernement tient-il les ...
BFMTV.COM - il y a 17 heures
Emmanuel **Macron** avait par ailleurs promis un "ministère plein et entier des Droits des femmes", qui a finalement été remplacé par un ...
DIAPORAMA - Les ministres du gouvernement **Macron**
RTL.fr - 17 mai 2017
Macron nommé un Premier **ministre** de droite, proche de Juppé
International - **Clicanoo** - 15 mai 2017
Edouard Philippe Premier **ministre** : comment Emmanuel **Macron** ...
Éditorial - **Challenges.fr** - 15 mai 2017
Qui sont les ministres du premier gouvernement **Macron**?
Exhaustif - **La Voix du Nord** - 17 mai 2017
Echos de campagne : Edouard PHILIPPE premier **ministre**
International - **Mayotte 1ère** - 15 mai 2017

Produit : image et texte

Lieu d'intérêt : image

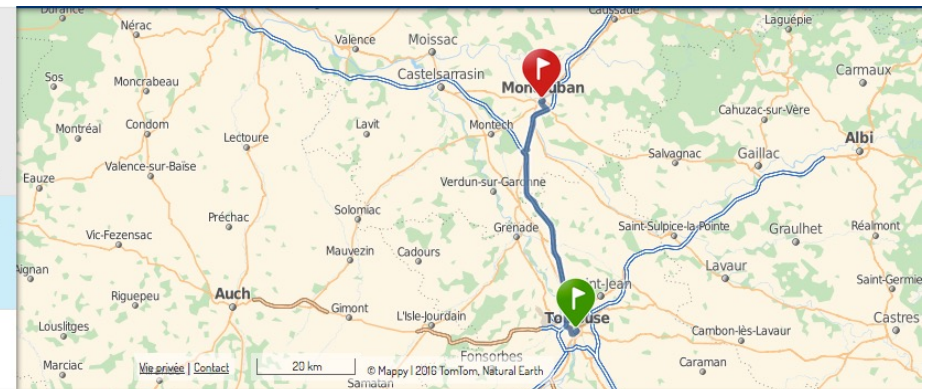


Toulouse 31000-31500

Montauban 82000

Via A62 54 min
9,11 € - 55,7 km

Voir la feuille de route



Itinéraire : image et contenu structuré

Actualités

Bac 2018: les sujets de l'épreuve d'histoire-géographie
BFMTV.COM - il y a 30 minutes
Voir la couverture complète

Melania Trump, «good cop» face à la politique de son mari
Libération - aujourd'hui
Immigration aux Etats-Unis : la polémique enfle, Melania Trump sort de sa réserve
Le Monde - hier
Afficher plus

Et si le séminaire franco-allemand était le dernier de Merkel?
Le HuffPost - aujourd'hui
« Je ne peux plus travailler avec cette femme » : Horst Seehofer lance un ultimatum à Angela Merkel
RT en français - aujourd'hui
Afficher plus

L'addiction aux jeux vidéo est désormais reconnue comme une maladie par l'OMS
Le Figaro - aujourd'hui
L'addiction aux jeux vidéo reconnue comme maladie par l'OMS
L'Usion - il y a 3 heures
Afficher plus

Pour vous
Recommandé en fonction de vos centres d'intérêt
Plus de contenus : Pour vous

Quand Cavani annonce l'arrivée de Ronaldo au PSG
FootBall.fr - aujourd'hui
Mercato - PSG : Nouveau coup de tonnerre dans le dossier Cristiano Ronaldo ?
Le 10 Sport - aujourd'hui
Afficher plus

Cinq erreurs de français et d'orthographe à éviter dans vos mails
Le Figaro - il y a 3 heures

Au sud-est de Toulouse, ce petit quartier va muter pour pouvoir accueillir 15 000 nouveaux habitants
actu - il y a 3 heures

La plus jeune, le surdoué recalé de Parcoursup, la championne de natation... Ces candidats remarquables du ...

C | F | K Lire la suite sur weather.com

Fact-checking
Non, on ne peut pas dire qu'il y aura deux millions d'immigrants régularisés en plus en France dans les dix prochaines années
Francinfo

Le vrai du faux. Non, ces vidéos ne montrent pas "l'orgie sexuelle" des joueurs mexicains
Francinfo

L'Afrique importe-t-elle 24 millions de tonnes de riz par an ?
africa check

Gros plan
BFMTV. « Les Hallyday n'étaient pas une vraie famille », selon une amie de Laura Smet
Ouest-France

Pourquoi doit-on se méfier des clés USB ?
Le Figaro

Le gouvernement coupe l'herbe sous le pied des coffee-shops
Libération

Coupe du monde : qui sont les favoris ?
Europe1

Ces vingténaires qui divorcent
Le Figaro
« Pognon de dingue » ? Lettre d'un allocataire du RSA au président Macron
<http://www.revolutionpermanente.fr/Section-Politique>
(Communiqué de presse)

Pourquoi les moustiques piquent-ils plus certaines personnes ?
Ouest-France

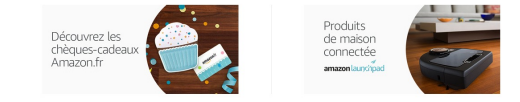
Ma "méthode voyoute" pour viser le 15/20 au bac de philo en manipulant votre correcteur

Personnes à suivre

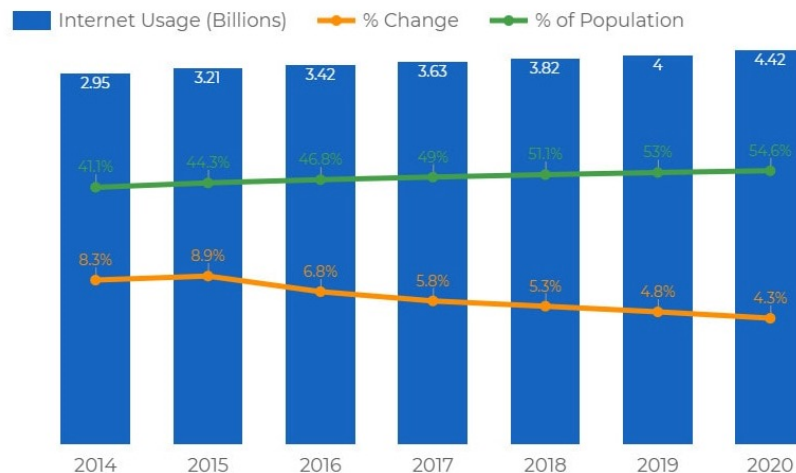
Recherchez un nom complet ou un @pseudonyme Recherchez sur Twitter

- Andrew Ng** @AndrewYNg
Co-Founder of Coursera; Stanford CS adjunct faculty. Former head of Baidu AI Group/Google Brain. #ai #machinelearning, #deeplearning #MOOCs
Suivre
- Liadh Kelly** @liadhkelly
Suivi par Jaime Teevan.
Suivre
- Fei-Fei Li** @drfeifei
Prof (CS @Stanford), Director (Stanford AI Lab), Chief Scientist AI/ML (Google Cloud), CoFounder/Chair @AI4allorg, Researcher #AI #computervision #ML cogneuro
Suivre
- rafaele perego** @rafaele_perego
computer scientist
Suivi par Fabrizio Silvestri, Guido Zucco et Jaime Teevan.
Suivre
- Andrej Karpathy** @karpathy
Director of AI at Tesla. Previously a Research Scientist at OpenAI, and CS PhD student at Stanford. I like to train Deep Neural Nets on large datasets.
Suivre
- Demis Hassabis** @demishassabis
Founder & CEO DeepMind - developers of #AlphaGo, #AlphaZero & Atari DQN. Working on General AI. Trying to understand what is "really" going on in the universe
Suivre
- Sérgio Nunes** @ssn
Web, Information, Media
Suivi par Fabrizio Silvestri et Miles Efron.
Suivre
- Dyaa Albakour** @dyaaa
Applied researcher in Information Retrieval and Text Analytics. Watch this space mainly for updates on events I attend or general posts on data science.
Suivi par Fabrizio Silvestri, SIGIR Forum et Krisztian Balog.
Suivre
- Riteish Deshmukh** @Riteishd
Actor, Architect, Producer : Mumbai Film Company
@mfc
Suivre
- CNN Breaking News** @cnbrk
Breaking news from CNN Digital. Now 55M strong. Check @cnn for all things CNN, breaking and more. Download the app for custom alerts: cnn.it/Apps
Suivre
- Akshay Kumar** @akshaykumar
Suivre

Produits



- Google traite en 2020 plus de **7 milliards de requêtes de recherche chaque jour** dans le monde 15% de ces requêtes n'ont jamais été recherchées sur Google auparavant
- On estime qu'il y aura environ **2.77 milliards d'utilisateurs de médias sociaux** dans le monde, contre 2.46 milliards en 2017
- **51%, ou plus de 240 milliards de dollars, de tout l'argent publicitaire** dépensé dans le monde en 2019 seront basés sur les médias numériques.
- Les ventes en ligne devraient atteindre **3.45 billions de dollars** de ventes en 2020
- 47.3% de la population mondiale devrait acheter en ligne en 2020.



Statistiques d'utilisation des applications d'accès à l'information 2014-2020

Source :

https://datastudio.google.com/embed/reporting/1sImC_rjeWqNXdgQt5MtmrQMbH44qFjtA/page/1fzh



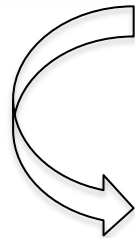
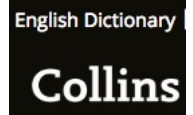
Solution technologique : **Personnaliser (automatiquement) l'information délivrée à l'utilisateur**

"Personalization, broadly known as **customization**, consists of tailoring a service or a product to accommodate **specific individuals**, sometimes tied to groups or segments of individuals.

A wide variety of organizations use personalization to improve **customer satisfaction, digital sales conversion, marketing results, branding, and improved website metrics as well as for advertising"**



"If you personalize something, you do or design it specially according to the needs of an individual or to your own needs"



L'individu, utilisateur, consommateur au centre du processus de personnalisation automatique de l'information

De très très...nombreuses applications

Achat en ligne (Amazon)

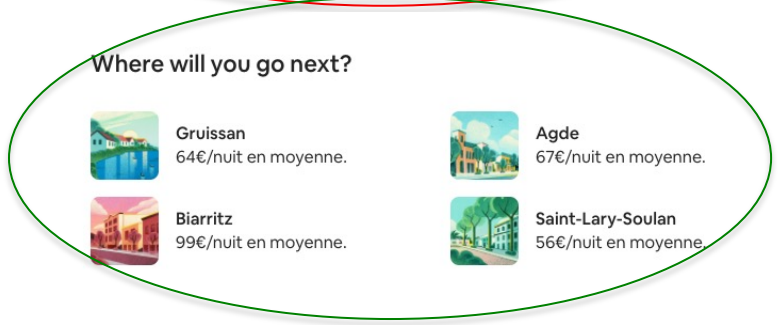
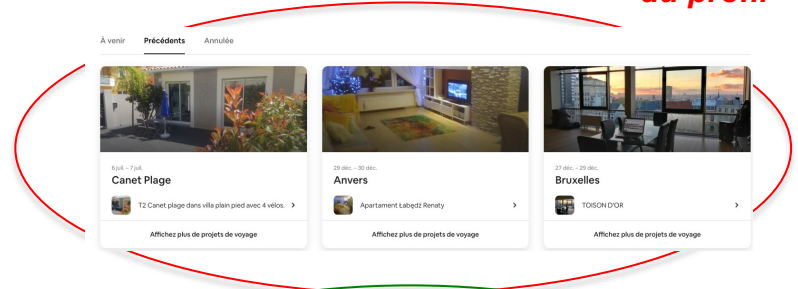
Partie des données du profil



Produits recommandés

Achat en ligne (AirBnb)

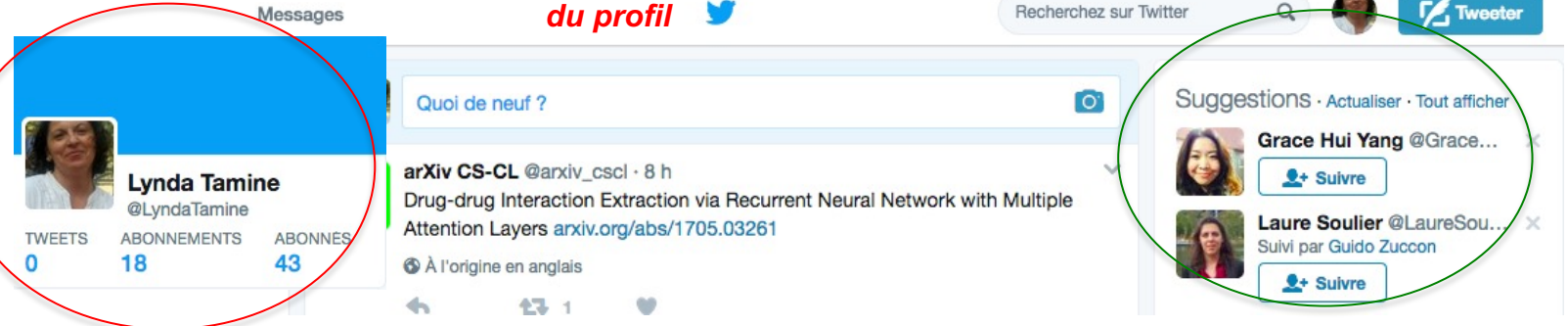
Partie des données du profil



Lieux recommandés

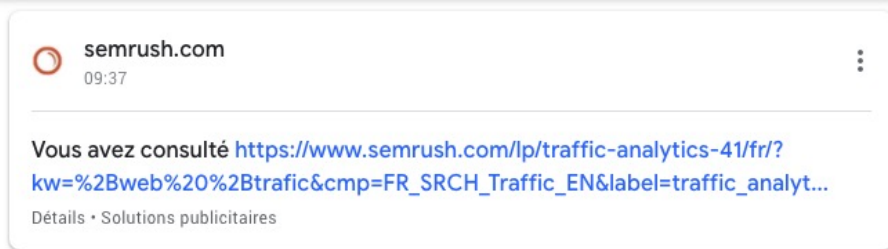
Plateformes sociales(Twitter)


Partie des données du profil




Personnes recommandées

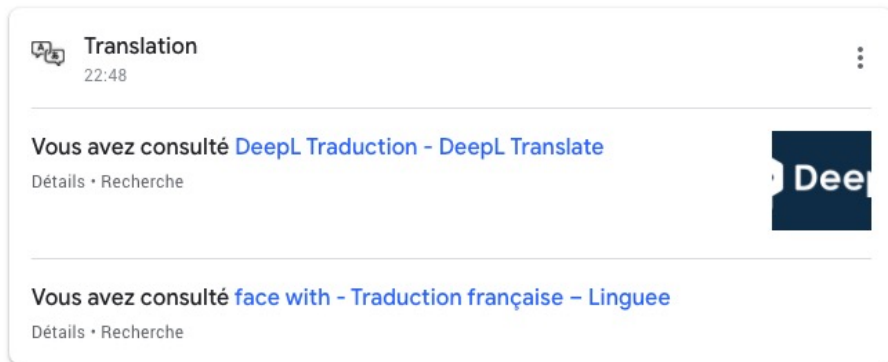
De très très...nombreuses applications





 **semrush.com**
09:37

Vous avez consulté https://www.semrush.com/lp/traffic-analytics-41/fr/?kw=%2Bweb%20%2Btrafic&cmp=FR_SRCH_Traffic_EN&label=traffic_analyt...
Détails • Solutions publicitaires

Hier 



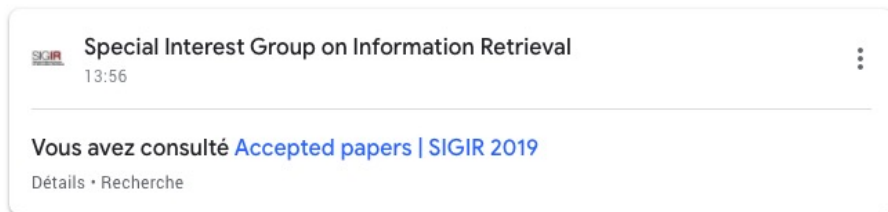
 **Translation**
22:48


Vous avez consulté [DeepL Traduction - DeepL Translate](#) 
Détails • Recherche

Vous avez consulté [face with - Traduction française - Linguee](#)
Détails • Recherche



 **google.com** (3 fois)
22:48



 **Special Interest Group on Information Retrieval**
13:56

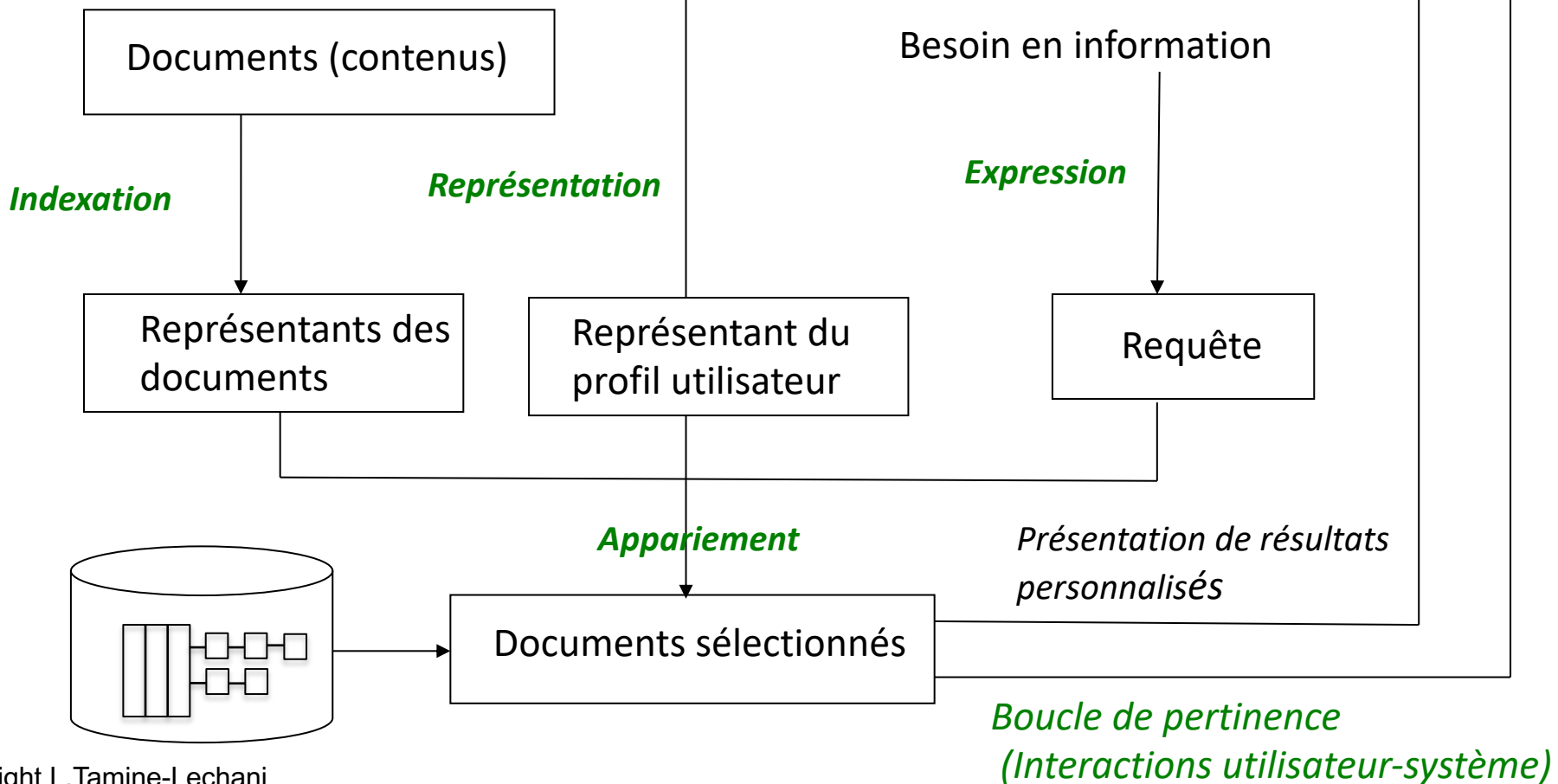
Vous avez consulté [Accepted papers | SIGIR 2019](#)
Détails • Recherche

☰ **Google** Mon activité

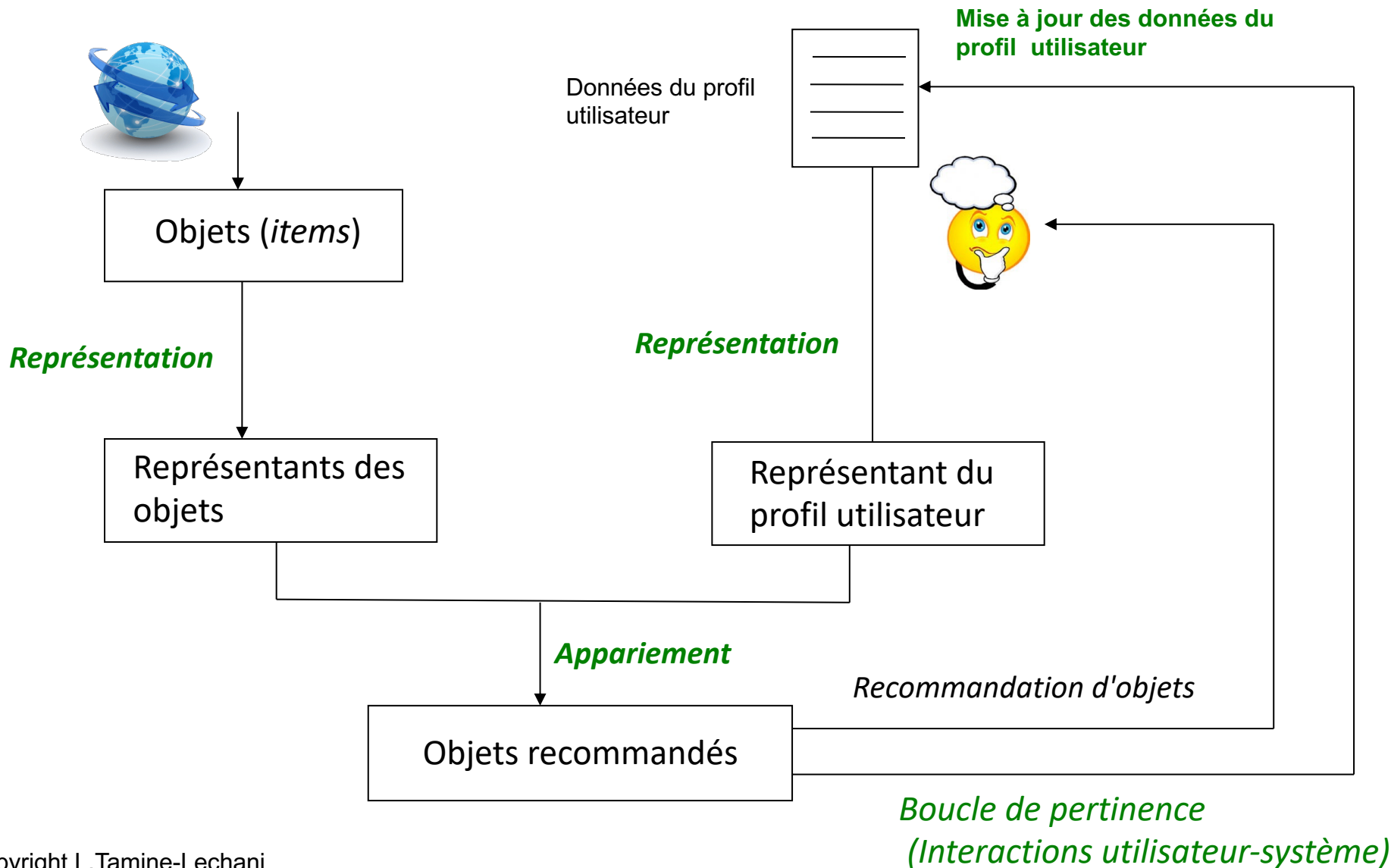
Moteur de recherche (Google)

*Partie des données
du profil*

☐ Systèmes de Recherche d'Information Personnalisée (SRIP)



❑ Systèmes de Recommandation (SR)



Recommandation et Personnalisation de l'Information

Plan du cours

Introduction

- Recommandation et accès personnalisé à l'information : objectifs
- Notion de profil/modèle utilisateur

Partie 1 : Systèmes de recherche d'information personnalisés

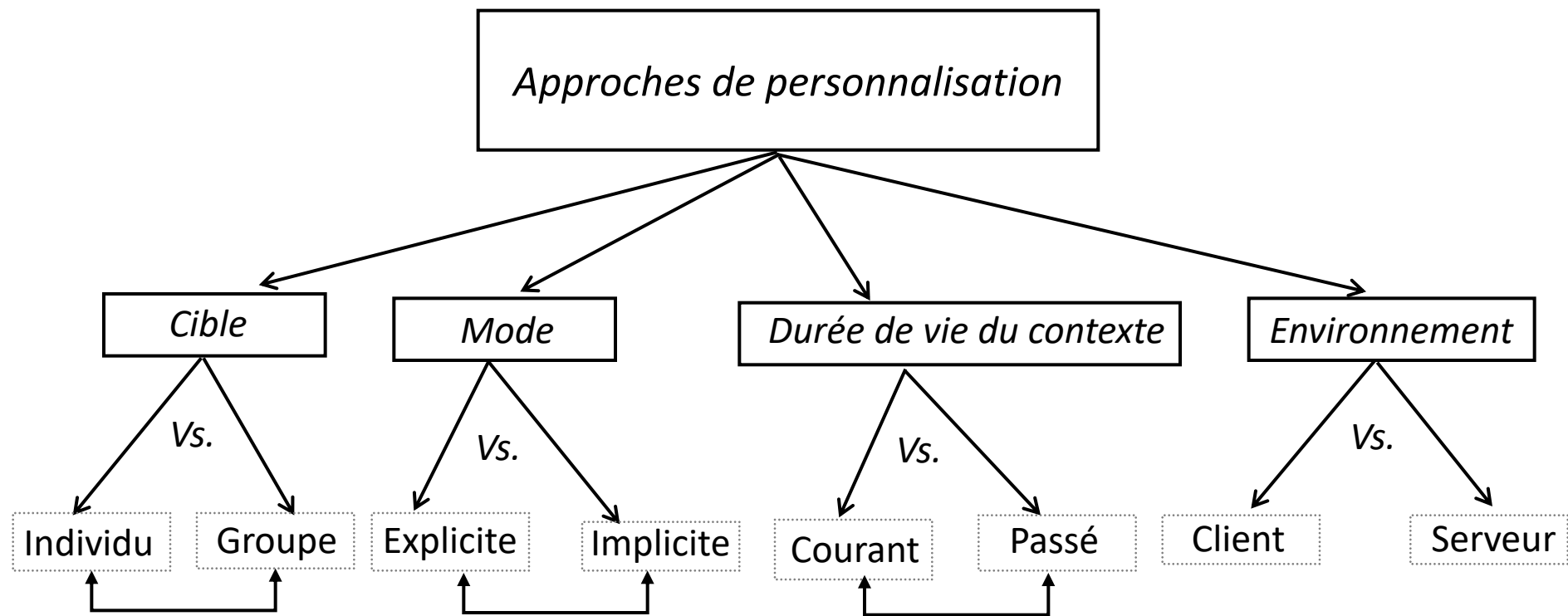
1. Modèles de représentation des profils utilisateurs
2. Modèles de recherche d'information personnalisée
3. Evaluation des systèmes de recherche d'information personnalisée

Partie 2 : Protection de la vie privée dans les systèmes de recommandation et systèmes de recherche d'information personnalisée

1. Données sensibles et risques dans les systèmes de recommandation et systèmes de recherche d'information personnalisée
2. Méthodes protection de la vie privée

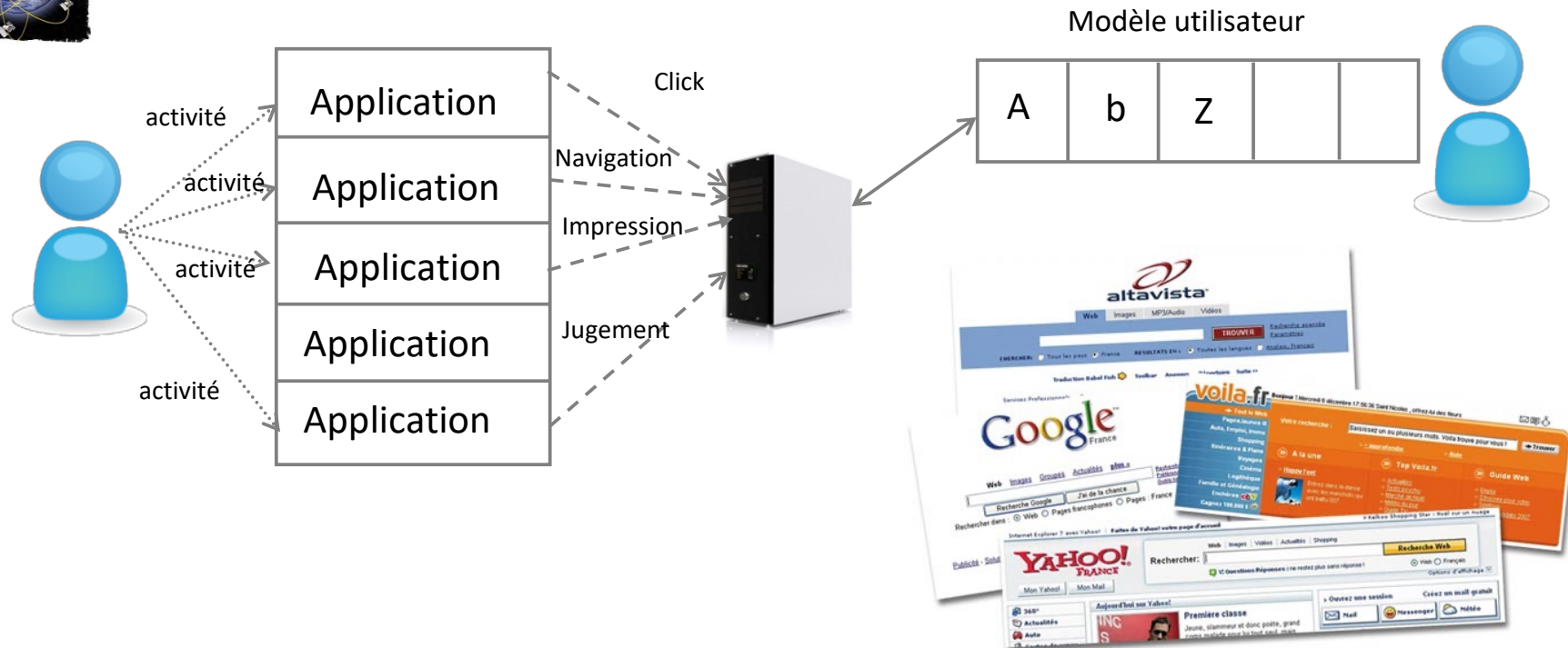
Partie 3 : Systèmes de recommandation

1. Modèles de recommandation
2. Evaluation des systèmes de recommandation



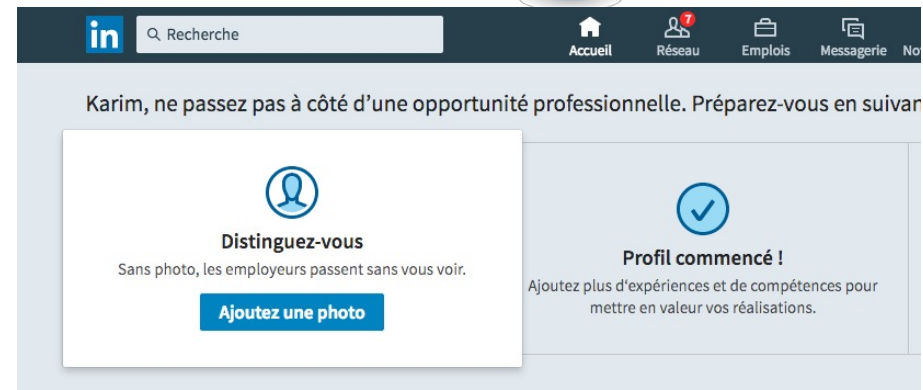
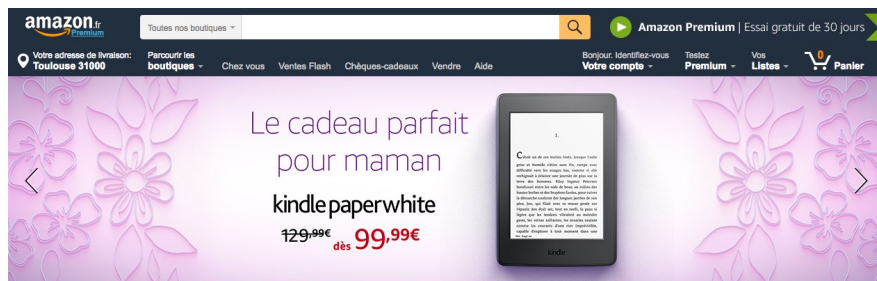
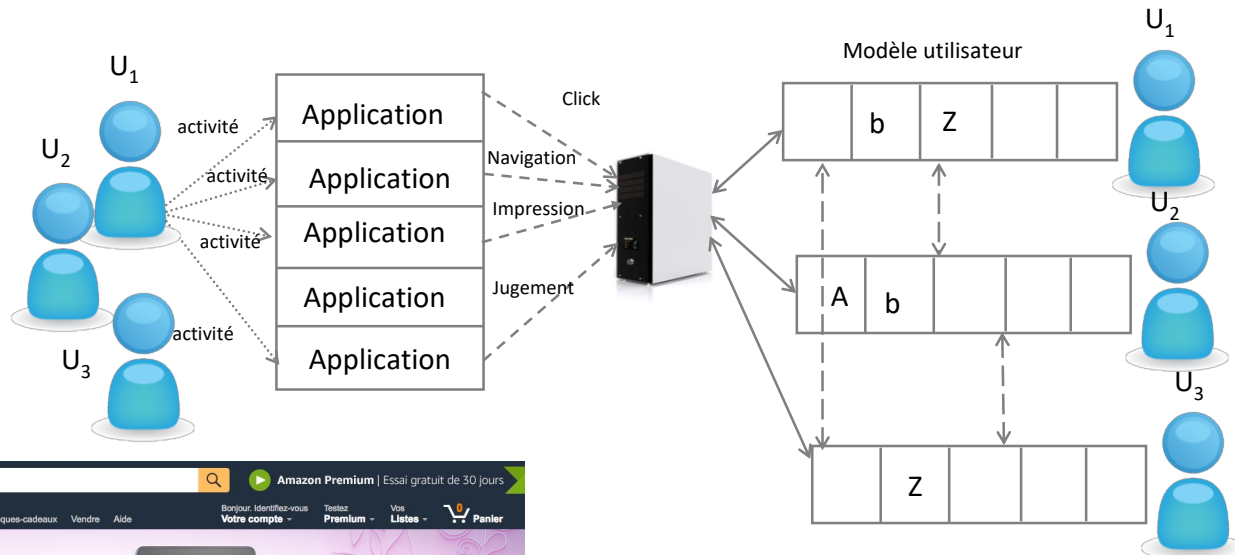
□ Personnalisation individuelle : principes

- ✓ Utilisateur considéré individuellement
- ✓ Profil/modèle utilisateur construit sur la base de ses préférences de contenu ou activités et mobilité via les applications



□ Personnalisation collaborative : principes

- ✓ Utilisateur considéré comme membre d'un groupe d'utilisateurs
- ✓ Modèle utilisateur construit sur la base de ses préférences de contenu ou activités via les applications, modèles des autres utilisateurs "similaires"



❑ Personnalisation explicite : données

- ✓ Collecte des données descriptives de l'utilisateur à partir de l'utilisateur lui-même
 - o Données démographiques (âge, sexe, langue maternelle etc...)
 - o Choix des préférences (langue, genre de documents, sujets d'intérêt)
 - o Jugements qualitatifs sur l'information (*'j'aime', 'je n'aime pas'*)

- ✓ Personnalisation explicite : sources et moyens de collecte des données
 - o Formulaires : cases à cocher, saisie de mots clés
 - o Interfaces élaborées : expression d'exemples, contre exemples, votes, notes, annotations etc.
 - o Questionnaires

❑ Personnalisation implicite : données

- ✓ Dérivation des données de profils à partir d'autres données disponibles
 - Données démographiques : peuvent être dérivées
 - ❖ des interactions et activités des utilisateurs
 - ❖ des requêtes
 - Données situationnelles
 - ❖ Lieu et temps d'émission de la requête
 - ❖ Environnement social : amis ou personnes physiquement proches
 - ❖ Agenda électronique
 - Données sociales
 - ❖ Annotations, posts, blogs etc.

☐ Personnalisation implicite : données

○ Activités de l'utilisateur

- ❖ Pages visitées, requêtes passées
- ❖ Données de navigation
- ❖ Applications utilisées
- ❖ Favoris
- ❖ Historique de localisations
- ❖ Interactions de l' utilisateur
 - ✧ Mouvements des yeux
 - ✧ Données de clicks
 - ✧ Actions sur les documents (ouverture, fermeture, impression, temps de lecture etc.)
 - ✧ Messages (e-mails) envoyés ou reçus

☐ Personnalisation implicite : sources et moyens de collecte des données

✓ Collecte des données descriptives de l'utilisateur à partir de ses activités, à travers :

o Le navigateur ou proxy

☺ Partage des données sur plusieurs sites (appli Web)

☹ Données récupérées que d'une seule machine

☹ Partage de données machine-serveur non facile

o Applications clientes (barre de bureau *toolbar*)

☺ Partage des données sur plusieurs sites (appli Web)

☺ Accès au bureau, fichier système etc.

☹ Nécessité de l'installation de l'agent (*toolbar*)

☹ Vulnérabilité des accès agent-spiroware

o Compte utilisateur (cookies, données de *log*)

☺ Historique de l'ensemble des activités

☹ Nécessité de créer le compte

☐ Personnalisation implicite : sources et moyens de collecte des données

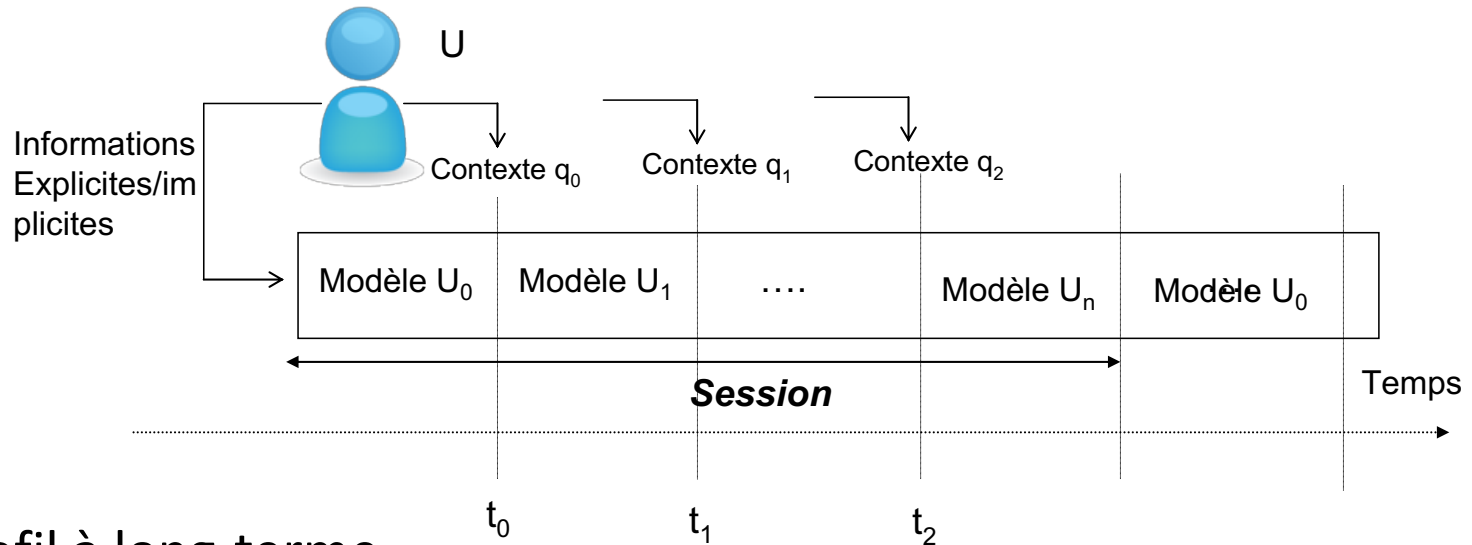
- ✓ Collecte des données descriptives du contexte spatio-temporel de l'utilisateur à travers son infrastructure matérielle :
 - o Global Positioning System (GPS)
 - ☺ Données précises sur le lieu, instant, orientation
 - ☹ Fiable pour le positionnement général, pas local
 - o Systèmes de positionnement infrarouge
 - o Réseaux sans fil
 - o Technologie *Bluetooth*
 - o *Horloge interne*

❑ Personnalisation hybride : combine la personnalisation explicite et la personnalisation implicite

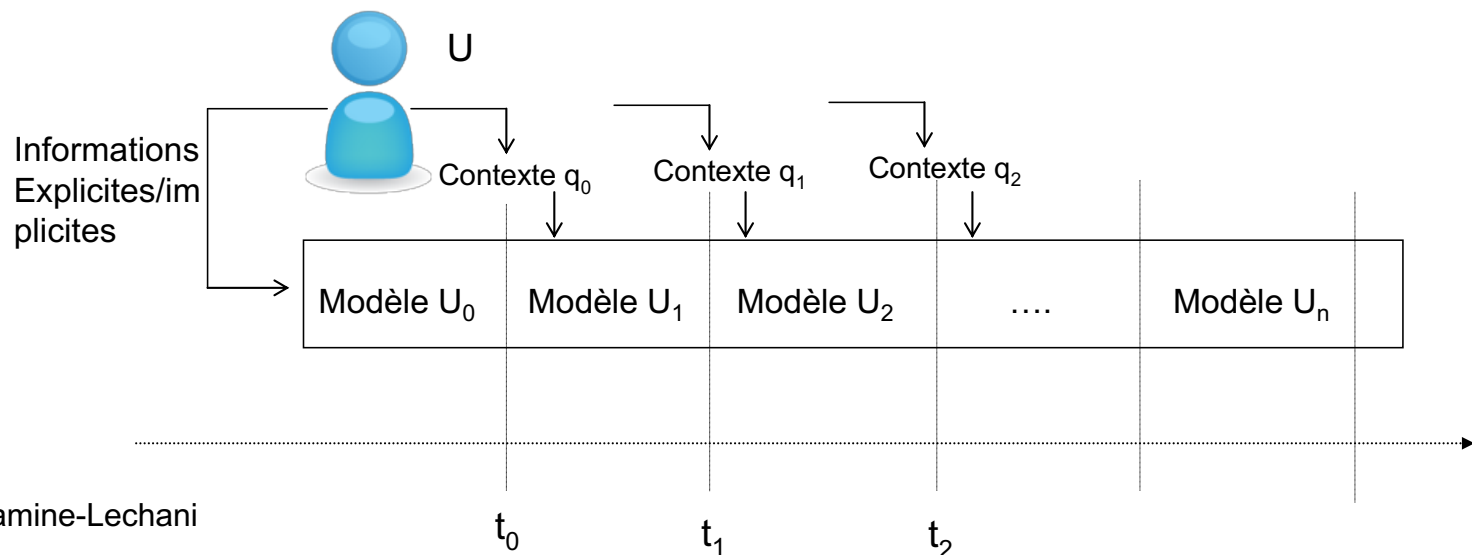
- ✓ Partie données du profil collectées explicitement à partir de l'utilisateur
- ✓ Partie des données du profil inférées à partir des données explicites

	Avantages	Inconvénients
Personnalisation explicite	<ul style="list-style-type: none"> - Facile à implémenter - Profil plus ou moins précis 	<ul style="list-style-type: none"> - Surcharge cognitive (temps, effort) - Qualité dépend de l'utilisateur (données erronées volontairement ou accidentellement) - Profil non mis à jour dynamiquement
Personnalisation implicite	<ul style="list-style-type: none"> - Pas de surcharge cognitive - Profil plus ou moins précis 	<ul style="list-style-type: none"> - Nécessite une technique d'inférence de données

□ Profil à court-terme

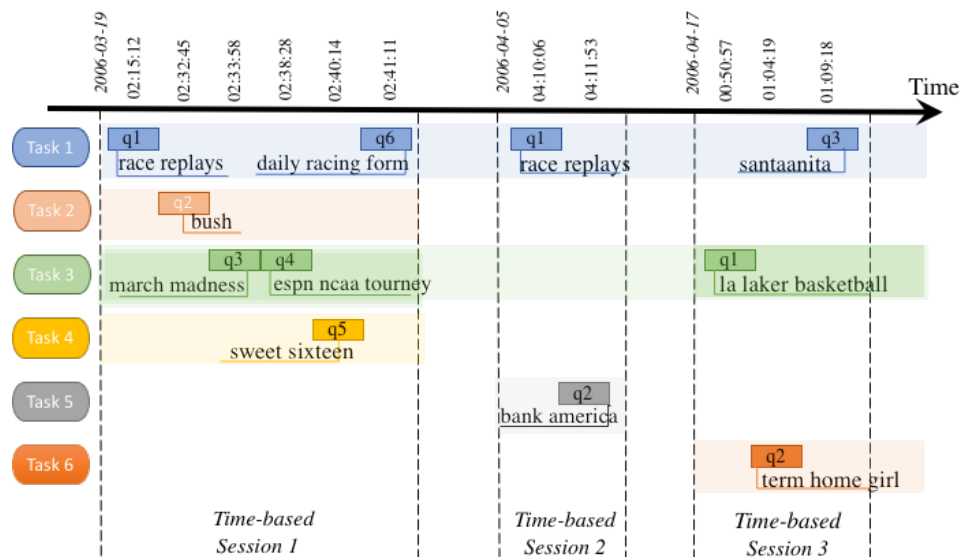


□ Profil à long-terme



❑ Personnalisation à court-terme : utilisation du **profil à court-terme**

- ✓ Mémoire de personnalisation = durée de la session
- ✓ Problème de délimitation de la session de recherche
 - o *Session physique* : délimitation basée sur le temps (30 min, 90 min d'inactivité entre 2 requêtes successives)
 - o *Session tâche* : délimitation basée sur la similitude des requêtes, résultats de recherche



❑ Personnalisation à long-terme : utilisation du **profil à long-terme**

- ✓ Mémoire de personnalisation = durée du compte utilisateur
- ✓ Principe d'ajout (accumulation des données), évolution des centres d'intérêts
- ✓ Délimitation des sessions de recherche ou sans délimitation

❑ Comparaison des système selon l'environnement source pour la personnalisation

	Personnalisation niveau client	Personnalisation niveau serveur
Données accessibles	<ul style="list-style-type: none"> - Données utilisateur de la machine cliente - Partage des données sur différentes applications 	<ul style="list-style-type: none"> - Données de tous les visiteurs du site hébergé par le serveur - Données non partagées sur différentes applications
Cible	<ul style="list-style-type: none"> - Personnalisation individuelle - Personnalisation collaborative non applicable 	<ul style="list-style-type: none"> - Personnalisation individuelle et personnalisation collaborative applicables
Mode	<ul style="list-style-type: none"> - Implicite - Explicite 	<ul style="list-style-type: none"> - Implicite - Explicite
Durée de vie	<ul style="list-style-type: none"> - Contexte courant - Contexte passé 	<ul style="list-style-type: none"> - Contexte courant - Contexte passé

- ❑ Le profil de l'utilisateur peut comporter divers types de données liées à :
 - ✓ **Personne** : âge, sexe, favoris etc.
 - ✓ **Comportement** : activités avec le système (navigation, annotations, feedback, etc.)
 - ✓ **Localisation** : lieu courant, historique de mobilité
 - ✓ **Voisinage social** : amis, abonnés, etc.
 - ✓ ...
- ❑ Le profil de l'utilisateur est construit
 - ✓ De façon **explicite** : données fournies par l'utilisateur
 - ✓ De façon **implicite** : données inférées à partir d'autres données
- ❑ Le profil utilisateur est exploité pour :
 - ✓ **Recommander à l'utilisateur** des informations **adaptés à son profil à court-terme ou profil à long-terme**
 - ✓ **Sélectionner** des informations **pertinentes à une requête** exprimée par l'utilisateur, **adaptées à son profil à court-terme ou profil à long-terme**