

Les données que récolte Google - Ch. 2

Voici déjà la traduction du deuxième chapitre de Google Data Collection, l'étude élaborée par l'équipe du professeur Douglas C. Schmidt, spécialiste des systèmes logiciels, chercheur et enseignant à l'Université Vanderbilt. Si vous les avez manqués, retrouvez les chapitres précédents déjà publiés.

Il s'agit aujourd'hui d'une expérience d'usage quotidien ordinaire du numérique en milieu urbain et connecté, expérience qu'il n'est pas trop difficile de transposer de ce côté-ci de l'Atlantique, et qui permet de repérer les différentes sortes de collecte opérées par Google, directement ou non.

Traduction Framalang : goofy, Khrys, serici. Remerciements particuliers à badumtss qui a contribué à la traduction de l'infographie.

II. Une journée dans la vie d'une utilisatrice de Google



« Earbuds », photo de susanjanegolding (CC BY 2.0)

5.

Afin d'illustrer la multitude des interactions entre Google et un individu, ainsi que l'étendue des informations collectées lors de ces interactions, nous avons réalisé une expérience dans laquelle un chercheur utilise un périphérique Android¹ pendant les activités d'une journée.

Afin d'éviter que des informations d'un utilisateur précédent ne soient associées au téléphone mobile, celui-ci a été réinitialisé aux valeurs d'usine² et configuré comme un téléphone neuf³. Un nouveau compte Google a été créé (nom d'utilisatrice : « Jane »), afin que Google n'ait pas de connaissances antérieures sur cette utilisatrice et qu'il n'ait pas associé de centres d'intérêts publicitaires à son compte. Le chercheur a passé une journée normale en utilisant son téléphone avec son nouveau compte Google.

6.

Les données collectées par Google ont été relevées par deux outils fournis par Google : « Mon activité »⁴ et « Télécharger vos données »⁵. L'outil « Mon activité » montre les données collectées par Google grâce à toute activité liée aux recherches, lors de l'utilisation des applications Google (i.e Youtube, Google Maps, Google assistant), par les visites sur des pages web tierces (lorsqu'on est connecté à Chrome), et les clics sur les publicités. L'outil « Télécharger vos données » fournit une information structurée concernant l'historique de toutes les données collectées par les applications Google (i.e cela contient tous les anciens courriels sur Gmail, toutes les recherches, l'ensemble des localisations et les vidéos YouTube consultées). Nous avons synthétisé les données collectées et nous les avons utilisées pour représenter les informations sur les événements clés dans l'infographie ci-dessous : « Un jour dans la vie » de l'utilisateur "Jane".

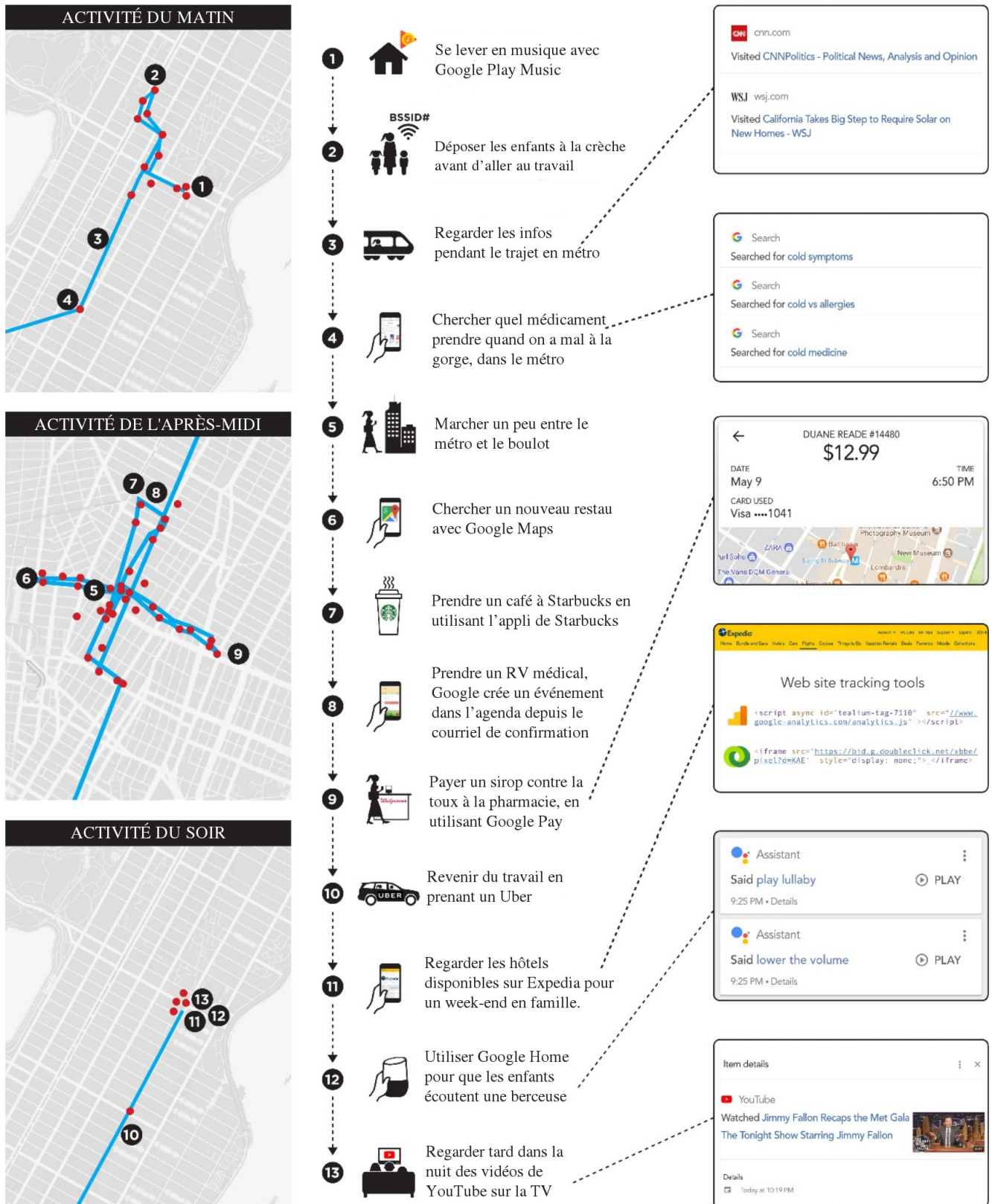


Figure 1 : une journée dans la vie d'une utilisatrice de Google

7. Pour l'activité détaillée dans la figure 1, mais également dans le reste du document, les données collectées sont classées en deux grands groupes : actives

et passives. Les données actives sont des données échangées directement entre l'utilisateur et un produit Google, là où les données passives sont définies comme une information transmise en arrière-plan sans notification évidente pour l'utilisateur. Par exemple, une collecte active de données est déclenchée lorsque Jane saisit un mot-clé dans l'outil de recherche et que cette requête est enregistrée par Google. Un exemple de collecte de données passives est l'envoi de la localisation de Jane à Google suite à l'enregistrement d'une requête.

8.

L'analyse des points de contact clés durant une journée normale dans la vie de Jane suggère que le nombre de données passives transmises est deux fois plus grand que le nombre de données actives (une décomposition détaillée des caractéristiques des données actives et passives est fournie dans le tableau qui figure en appendice, page 37 du document original).

9.

Google analyse les données collectées pour déterminer les centres d'intérêt des utilisateurs et utilisatrices, qu'il utilise ensuite pour les cibler avec des publicités adaptées. Par exemple, Google fournit une liste d'intérêts qu'il a déduits de l'activité d'un utilisateur, que l'on peut consulter sur la page « Les sujets qui vous intéressent » de la page de « Personnalisation des annonces » de Google ⁶. La figure 2 ci-dessous montre la liste que Google a associée avec le compte de Jane après une journée d'activité. Au total, Google a attribué 18 centres d'intérêts à Jane, dont 8 (entourés par une bordure rouge) qui correspondent précisément aux utilisations et activités de Jane⁷

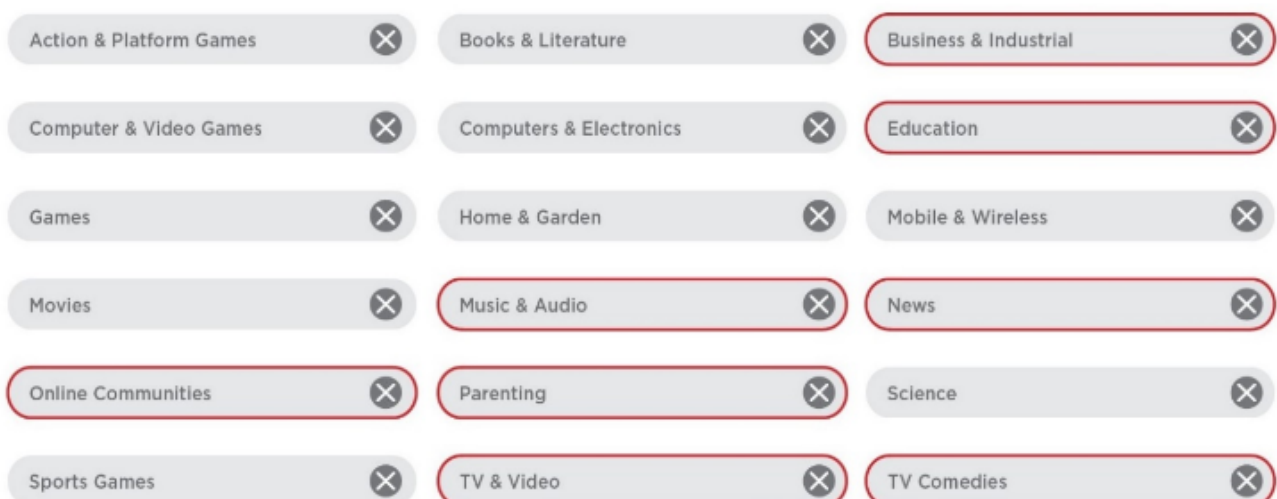


Figure 2 : les centres d'intérêt de Jane tels que les a déduits Google après une journée typique : réseaux sociaux, musique, parentalité, TV et vidéos, entreprise et industrie, éducation, nouvelles, comédies TV

10.

Bien que les outils « mon activité » et « Télécharger mes données » soient utiles pour estimer la quantité de données actives collectées lors des interactions d'un utilisateur avec les produits Google, ils ne dessinent pas une image complète de l'ampleur et de l'échelle de la collecte de données de Google. Comprendre cela requiert un passage en revue détaillé des clauses d'utilisation des produits en ce qui concerne la vie privée mais également l'analyse du trafic de données envoyé aux serveurs de Google pendant une session d'utilisation par un utilisateur de ces services. Les résultats de cette analyse sont présentés plus loin dans ce rapport.