L'AMITIÉ AU PREMIER REGARD.

DOCTORAT SAUVAGE EN NUMÉRIQUE

Échelle inconnue · Ventres mous

GOOGLE EST MON AMI

TOUS LES JOURS CHEZ VOUS

Mais aussi au boulot, en vacances, pendant les courses...

Google est mon ami

- Du Minitel à Google
  balbutiements de la recherche
- Les débuts de Google
  premiers défis
- Google aujourd'hui
  Monstre et compagnie
- Les services Google
  Nous sommes les frères qui rap'tout
- Aujourd'hui

Du Minitel à Google

Rechercher au temps du Minitel

- Annuaires papier Minitel
- Magazines et hors-séries
- Le Kiosque 3615
- Des outils de leur époque
  - navigation peu automatisable
  - pas d'hyperlien
  - nombre de services limités
  - indexation ➡ temps ➡ coût

Tarifs Minitel

		FF/heure	€/heure
Annuaire électronique	3611	22,20	5,08
Libre appel	3605	0,00	0,00
Télétel 1	3613	7,80	1,79
Télétel 2	3614	22,20	5,08
Télétel 3/3P	3615/3616	50,40	11,54
		58,80	13,46
		75,00	17,17
Télétel 4	3617/3621	131,40	30,08
Minitel Dialogue	3618	58,80	13,46
Minitel Direct	3621	7,80	1,79
ASCII	3628XYZT	328,50	75,21
Télétel	3629	543,50	124,43

Youpi, le web arrive!

- **1989** création du web
  Créé par Tim Berners-Lee au Cern
- **1993** premier web crawler (MIT)
  Son créateur, Matthew Gray, travaille depuis 2007 chez Google
- **1994** Yahoo!
  Répertoire de sites catalogués humainement

YAHOO!

La recherche devient essentielle

- **1994** Lycos
  - un des premiers moteurs de recherche
  - plus gros index en 1996 avec 60 millions de documents
- **1995** AltaVista
  - démonstration des ordinateurs Dec Alpha 8400 TurboLaser et de leur SGBD
  - 20 serveurs pour un total de 2,5 To RAM, 10 To disque

Des défis à relever 1/2

- Quantité énorme de documents
  à parcourir, à indexer, à stocker, à rechercher, à rafraîchir…
- Complexité des documents
  Langues, jeux de caractères, homonymes, fautes d'orthographe, formats…
- Pertinence des résultats
  Un même mot peut avoir une signification différente pour deux personnes
- Évolution des algorithmes et des normes

Des défis à relever 2/2

- Pannes, obsolescence et mises à jour
  Le service doit être disponible 24h/24, 365j/an
- Puissance de calcul et consommation électrique
- Connexion Internet
  L'ADSL arrive en France en 1999
- Tricheurs, sécurité
- Mais surtout… comment gagner de l'argent ?

Les débuts de Google

La rencontre

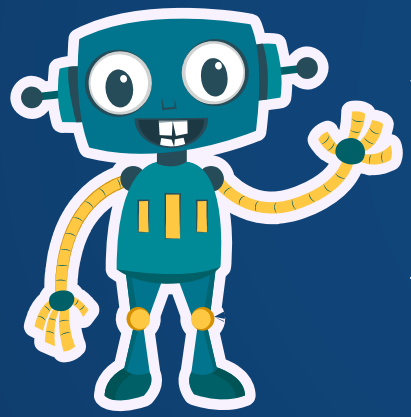
- **1995** Larry Page et Sergueï Brin se rencontrent
  Université de Stanford dans la Silicon Valley
- **1996** ils collaborent sur 3 projets
  - robot d'indexation
  - algorithme PageRank
  - moteur de recherche BackRub

Le lancement

- **1997** google.com est réservé
- **1998** installation dans un garage
  une reconstitution est visitable sur Street View
- **1999** décollage
  - **février** : 500 000 recherches/jour
  - **mars** : déménagement à Palo Alto
  - **juin** : 25 000 000 $ de capitaux propre
  - **août** : 3 000 000 recherches/jour
  - **septembre** : déménagement à Mountain View

https://www.youtube.com/watch?v=KoeXTH1_EN8
https://www.blog.google/products/maps/inside-googles-original-garage-1998-style/

À l’époque



A

B

C

	A	B	C
hibou	X	X	
joujou			X
caillou		X	X
genou		X	
chou	X		X

méta-données
date de collecte langue dernière mise à jour format en-têtes du serveur etc.

Indexation et pertinence

- Prise en compte du contexte
  ex.: « pied de biche » n'a aucun rapport avec l'animal
- Importance sémantique des mots
  ex.: le, la, les ont peu d'importance
- Glissement sémantique
  ex.: évolution du mot « typographie »
- Fautes d'orthographe
- Actualité ou information

PageRank

- Calcule la popularité d'une page
- Une page gagne en popularité si...
  - elle est pointée par une page populaire
  - la page la pointant n'a pas beaucoup d'autres liens
- Google a la plus grande popularité possible
- Un facteur parmi d'autres aujourd'hui
  Cette popularité peut être influencée par des fermes de sites

Mais...

- La pertinence peut encore être améliorée
- Reste le problème du modèle économique

$ € £ ¥...

Google aujourd'hui

Algorithmes, algorithmes

- Analyse fine des pages indexées
  Contextualisation du contenu, analyse de sa structure
- Personnalisation des résultats
  Une même recherche ne donne pas les mêmes résultats en fonction du lieu ou des centres d'intérêt de l'utilisateur
- Évolutions régulières
  Jeu du chat et de la souris de l'optimisation SEO, apparition de nouveaux mots, évolution de l'usage des mots et de leur signification

Un modèle économique

- Pour Google
  - ses clients : des annonceurs
  - leur produit : vous !
- Besoin de collecter un maximum de données
  - pour proposer le meilleur ciblage aux annonceurs
  - pour vous fournir le meilleur service
- Maximum de données ➡ maximum de traitements

15 datacenters

Quelques chiffres

- entre 900 000 et 2 000 000 de serveurs (estimations)
- 1 datacenter = plusieurs bâtiments
- 1 bâtiment = 75 000 serveurs
  100 To/s de bande passante
- 2,6 gigawatts d'énergie renouvelable
  Google consomme notamment la totalité des 114 méga-watts produits par une ferme d'éoliennes dans l'Iowa depuis 2010
- Google reste discret sur ses chiffres

https://www.youtube.com/watch?v=zDAYZU4A3w0
http://www.datacenterknowledge.com/archives/2017/03/16/google-data-center-faq

Les services Google

Les services Google

- Gratuité
- Simplicité
  Utilisation, installation, mise en œuvre…
- Performances
  Temps de réponse, bande passante, volumes…
- Disponibilité
- Facilement insérables dans une page web

Les données collectées

- **Ce que vous faites**
  recherches, sites web consultés, vidéos visionnées, annonces visitées, situation géographique, matériel, adresse IP, cookies
- **Ce que vous créez**
  e-mail via Gmail, contacts, agenda, photos, vidéos, Google Docs/Sheets/Slides, Drive
- **Ce qui vous rend unique**
  nom, adresse mail, mot de passe, date de naissance, sexe, numéro de téléphone, pays

https://privacy.google.com/intl/fr/your-data.html

Le Google visible

Google Search

- Moteur de recherche le plus rapide/pertinent
  - 30 000 milliards de pages indexées
  - 20 milliards de sites visités/jour
  - 3,3 milliards de requêtes/jour
- Monopole des recherches (94%)
  98% sur mobile, Qwant est à 0,58%
- Moteur de recherche par défaut
  Firefox, Chrome

https://www.blogdumoderateur.com/chiffres-google/
http://www.webrankinfo.com/dossiers/etudes/parts-marche-moteurs

Android

- Système d'exploitation pour mobiles et tablettes
- Monopole mondial (88%)
  1,6% en mars 2009
- Connexion maximale avec les services Google
  agenda, contacts, applications, drive, vidéo, géolocalisation...
- Google avec vous, tout le temps!

https://www.phonandroid.com/android-88-parts-marche-2017-graphique-revient-ascension-fulgurante.html

Android vous suit à la trace

- Même si la géolocalisation est désactivée
- Même s'il n'y a pas de carte Sim
- Même si aucune application n'est installée
- Google utilise tous les moyens possibles
  GPS, adresse IP, antennes des opérateurs, Wifi, Bluetooth

https://qz.com/1131515/google-collects-android-users-locations-even-when-location-services-are-disabled/

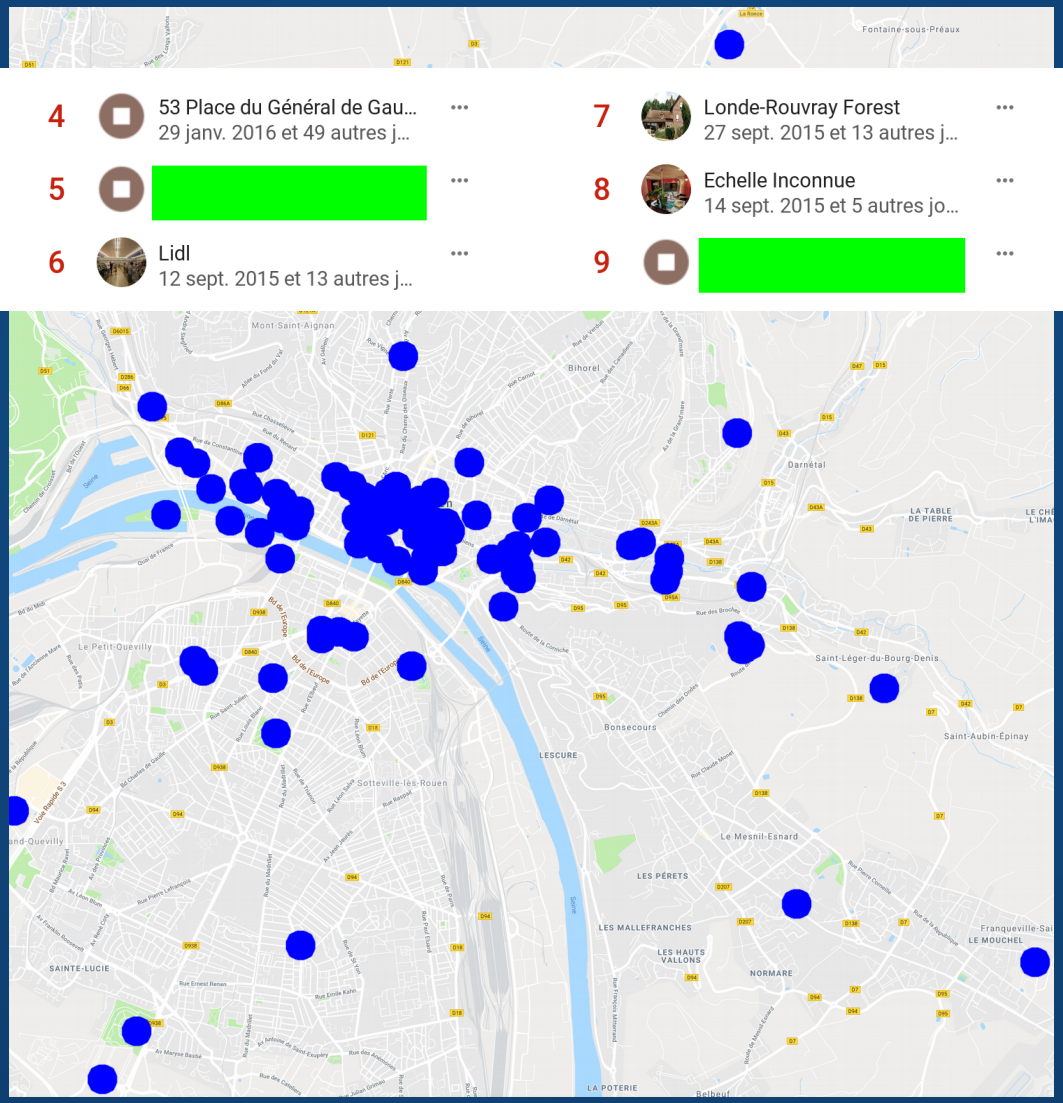
Google Maps

- Service de cartographie
- Rapidité, Street View, info trafic
- Où vous êtes, où vous allez, ce que vous recherchez
- Embarqué sur de nombreux sites

Android + Google Maps

- Où vous êtes
- Où vous avez été
- Quand vous y étiez
- Vos habitudes
  - les lieux que vous fréquentez le plus
  - combien de temps vous y restez
  - les enchaînements de lieux

Android + Google Maps



1 [redacted]

2 [redacted]

3 [redacted]

4 53 Place du Général de Gau...
29 janv. 2016 et 49 autres j...

5 [redacted]

6 Lidl
12 sept. 2015 et 13 autres j...

7 Londe-Rouvray Forest
27 sept. 2015 et 13 autres j...

8 Echelle Inconnue
14 sept. 2015 et 5 autres jo...

9 [redacted]

10 Musée National de l'Educat...
3 mai 2015 et 1 autre jour

11 Castorama Darnétal
18 sept. 2015 et 1 autre jour

12 [redacted]

https://www.google.com/maps/timeline

Gmail

- Messagerie électronique
- 15 Go d’espace
- Anti-spam performant
- Recherche made in Google
- Toutes vos correspondances lues par Google!
  projets d’achat ou d’investissement, mots de passe,
  mots plus ou moins doux, photos intimes...

Google Agenda

- Agenda électronique
  - vos rendez-vous
  - vos centres d’intérêt

Google Chrome

- Navigateur web
- Mode non connecté
  préférences linguistiques, assistance à la navigation, statistiques d'utilisation
- Mode connecté
  Synchronisation avec son compte Google (sites visités, favoris, mots de passe...)

https://www.google.com/intl/fr/chrome/browser/privacy/

YouTube

- Plateforme de partage de vidéos
- Les chiffres :
  - 30 millions d'utilisateurs/jour
  - 5 milliards de vidéos visionnées/jour dont 1 milliard sur mobile
  - 300 heures de vidéo ajoutées/minute
- Ce que vous aimez, détestez, partagez
- Embarqué sur de nombreux sites

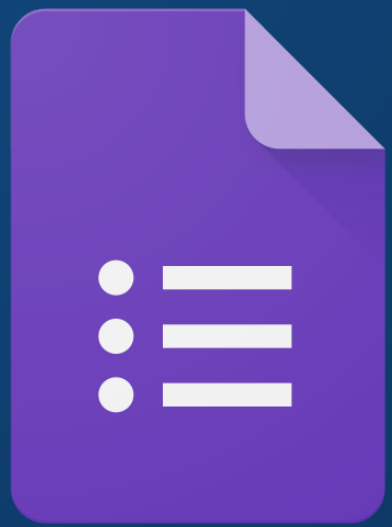
https://fortunelords.com/youtube-statistics/

Google Docs/Sheets/Slides

- Suite bureautique en ligne
- Édition simultanée de documents
- Documents soumis à la discrétion de Google…
- Google Drive

Google Forms

- Formulaires en ligne
- Collecte de données diverses



Google Assistant

- Assistant vocal bidirectionnel « Ok Google ! »
- Rendre instinctive l'utilisation de Google
- Écoute en permanence

Play Store

- Magasin d'applications Android
- Google sait quelles applications
  - vous installez
  - sur quels appareils
  - vos goûts
  - l'évolution de vos goûts

Google+

- Réseau social de Google
- Se ~~veut~~ voulait concurrent de Facebook
  Facebook limite fortement l'intrusion de Google
- Fermeture le 2 avril 2019
- Google n'est pas toujours gagnant

MEME QUAND VOUS N'ETES PLUS SUR GOOGLE VOUS ETES ENCORE SUR GOOGLE

Google Analytics

- Outil de statistiques pour site web
- Monopole en France (70% en 2014)
  sncf.com, metropole-rouen-normandie.fr, reseau-astuce.fr, onvasortir.com, rouentourisme.com, musees-rouen-normandie.fr...
- Installé par de nombreux prestataires
  très simple à installer et gratuit
- Son omniprésence permet à Google de suivre les internautes à la trace

http://www.journaldunet.com/solutions/analytics/marche-web-analytics.shtml

Google Fonts

- Plus de 900 fontes gratuites pour le web
  Roboto, Srisakdi, Open Sans, Lato, Raleway Poppins, Ubuntu, Lora...
- Hébergées par Google...
  Quand votre navigateur récupère la fonte chez Google, il donne l'adresse de la page où cette fonte est utilisée.

EXAMPLE

1
Salut example.com,
c'est le Firefox d'Émilie.
je voudrais la page
joli.html

2
Salut Firefox d'Émilie,
c'est example.com.
Voici la page joli.html

3
Le Firefox d'Émilie
m'a demandé
joli.html

4
La page a besoin de
la fonte Roboto
de Google

5
Salut Google
c'est le Firefox d'Émilie.
je voudrais la fonte
Roboto pour joli.html
de example.com

6
Salut Firefox d'Émilie,
c'est Google.
Voici la fonte Roboto

7
Le Firefox d'Émilie
a visité joli.html
sur example.com

Google

Hosted libraries

- Scripts utilisables par les développeurs
- Google se charge de l'hébergement
  Même remarque que pour les fontes
- Nombreux scripts disponibles
  D3 JS, Dojo, Ext Core, Hammer JS, Indefinite Observable, jQuery, jQuery Mobile, jQuery UI, MooTools, Myanmar Tools, Prototype, script.aculo.us, Shaka Player, SPF, SWFObject, three.js, Web Font Loader

https://developers.google.com/speed/libraries/

Google AMP

- Pages Mobiles Accélérées
- Accélère le chargement des pages
  Document simplifié, mise en cache
- Pages mises en avant par Google
- Pages servies sur les serveurs de Google

https://80x24.net/post/the-problem-with-amp/
https://www.ampproject.org/docs/guides/how_cached

reCAPTCHA

- Filtrage des robots
- Vous travaillez gratuitement pour Google!
  en entraînant leur intelligence artificielle

https://security.googleblog.com/2014/12/are-you-robot-introducing-no-captcha.html

Et encore...

Vous avez échappé à ça

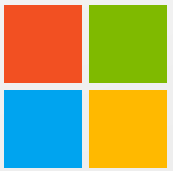
- Google Public DNS
- Google Trends
- Google AdSense
- Google AdWords
- Google APIs
- Google Cloud
- Pixel
- Google Photo
- Google Local Guide
- Google Earth
- Google Waze

Et aussi à ça

- Google ChromeCast
- Blogger
- Google Wifi
- Google Payments API
- Android Auto
- Google Play Musique
- Google SignIn
- Google Books
- Google Education
- Google Scholar
- ...

L'herbe est toujours
plus **verte** chez les autres
jusqu'à ce qu'on découvre
que c'est du **gazon artificiel**

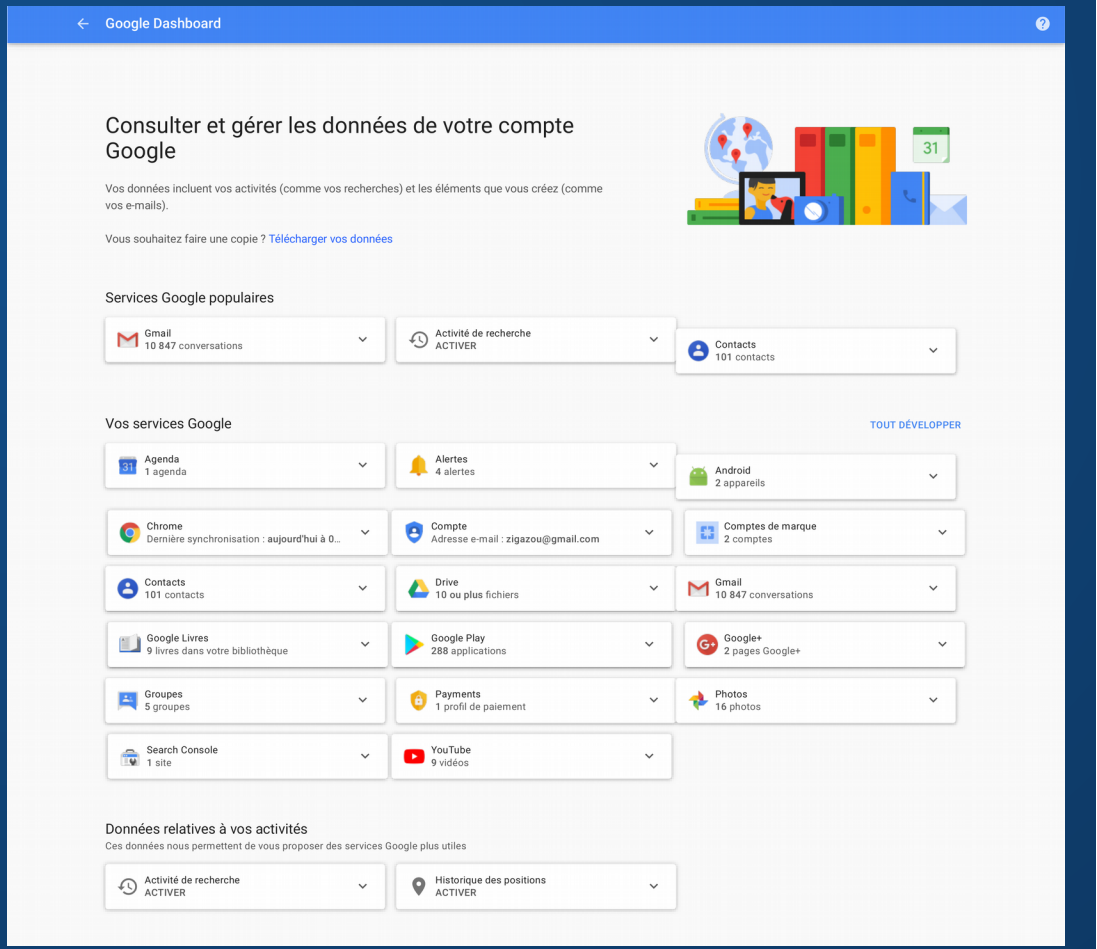
Jacques Salomé



Microsoft

Ce que Google sait de vous

- Google Dashboard
  https://myaccount.google.com/dashboard



Moteurs de recherche

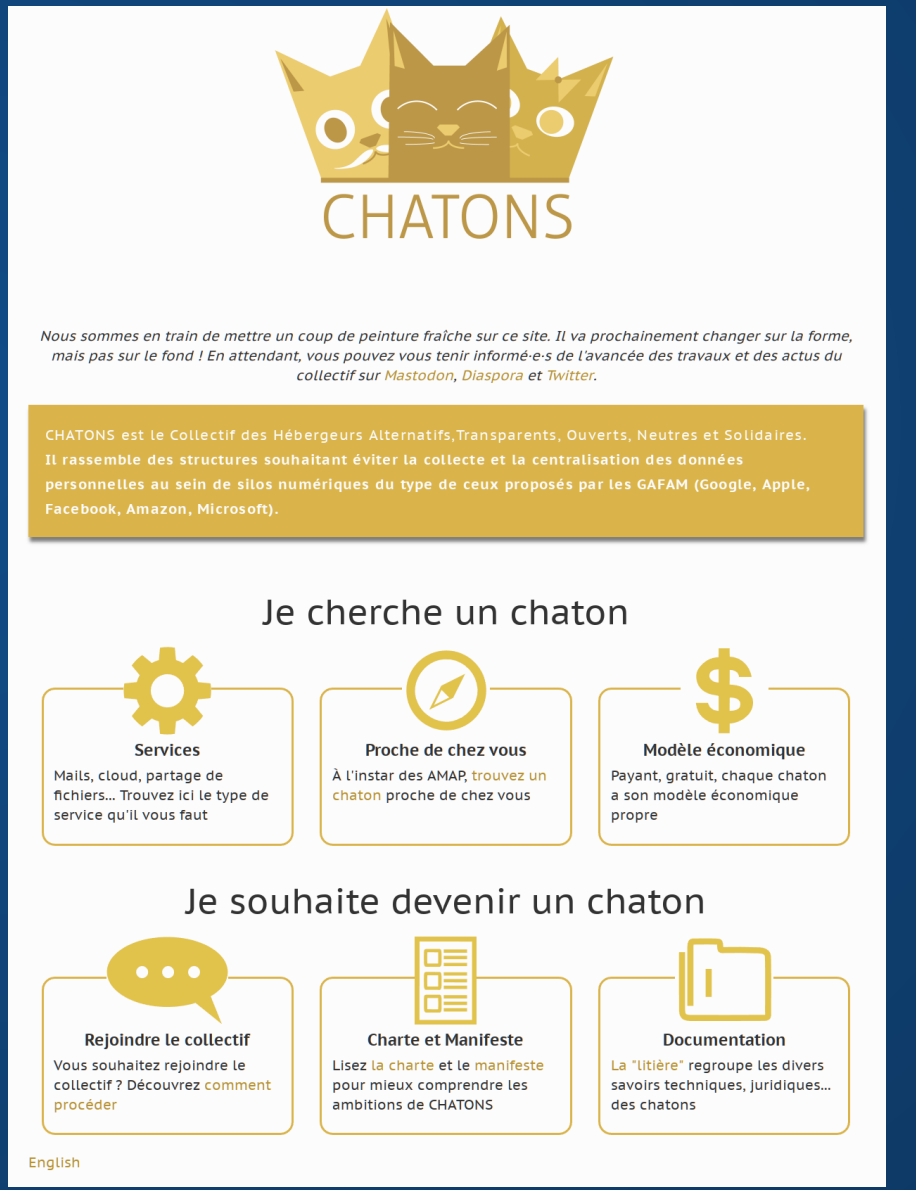
- Qwant
  https://www.qwant.com/
- DuckDuckGo
  https://duckduckgo.com/

Alternatives

- Dégooglisons Internet – FramaSoftt
  https://degooglisons-internet.org/

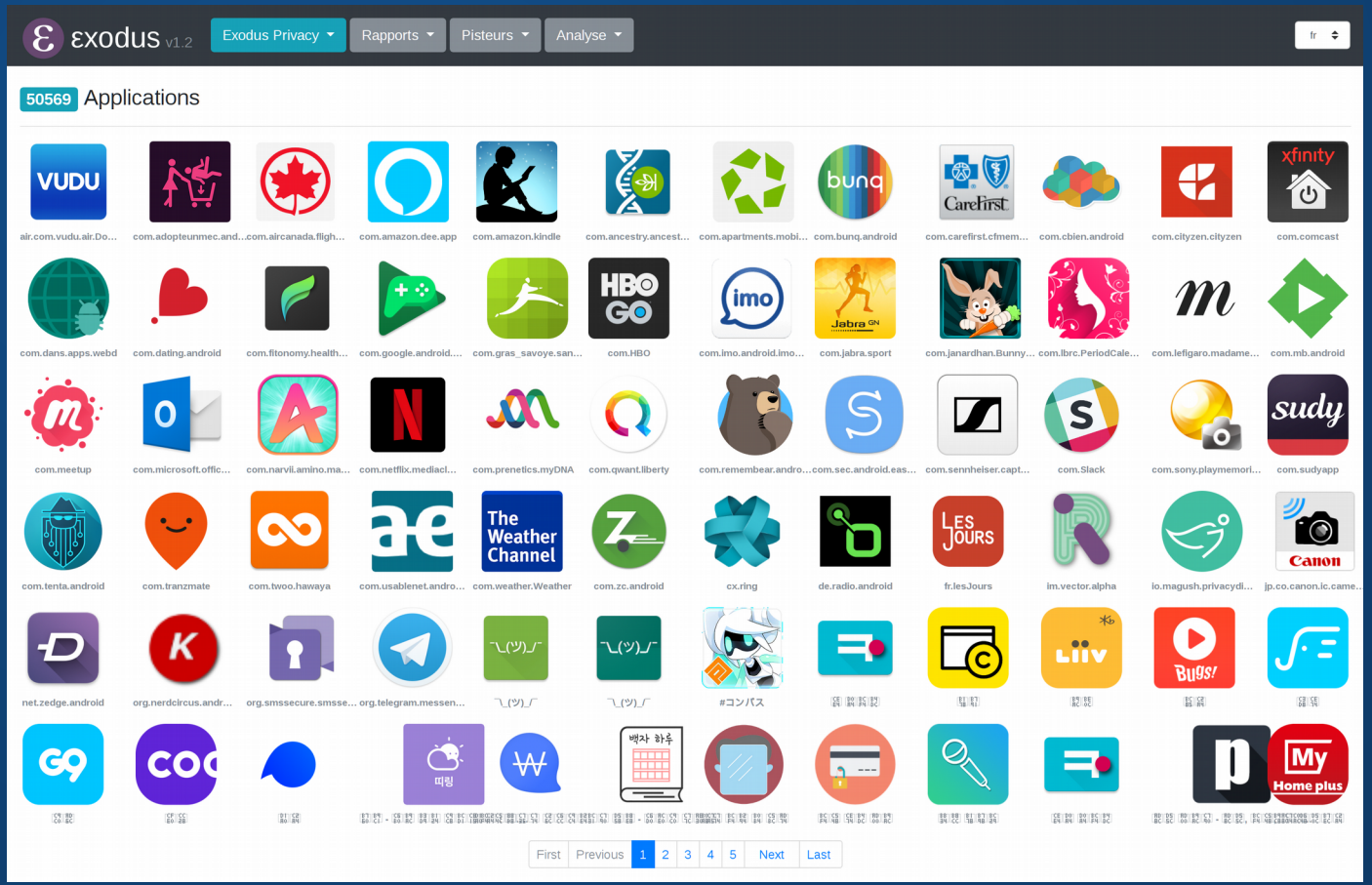
Bonnes âmes

- Chatons
  https://chatons.org/



Secrets peu avouables des apps

- Exodus Privacy
  https://exodus-privacy.eu.org/



4 pisteurs

13 permissions

RESEAU-ASTUCE

Version : **1.31**
Créateur : **Cityway Team**
Téléchargements : **50,000+ downloads**

Autres versions | Sur Google Play | Empreinte de la clé

Ce rapport a été automatiquement créé le **3 juin 2018 20:15**.
Ce rapport a été automatiquement mis à jour le **16 août 2018 15:59**.

Signé par :

Empreinte : ec211b2a528f1374d2d67c05e70e6998d9ea50d4
Émetteur : countryName=FR, stateOrProvinceName=Unknown, localityName=Unknown, organizationName=Cityway, organizationalUnitName=Unknown, commonName=Unknown
Sujet : countryName=FR, stateOrProvinceName=Unknown, localityName=Unknown, organizationName=Cityway, organizationalUnitName=Unknown, commonName=Unknown
Série : 1340031207

4 Pisteurs

Nous avons trouvé la **signature** des pisteurs suivants dans cette application :

- Google Ads
- Google Analytics
- Google DoubleClick
- Google Firebase Analytics

Un pisteur est une partie du logiciel dédiée à la collecte de données sur vous et vos usages. Nous ne garantissons pas l'exhaustivité de cette liste.

Voici la liste des signatures des pisteurs que nous avons trouvés par l'analyse statique de cet APK. Cela ne constitue pas une preuve de l'activité de ces pisteurs. L'application pourrait contenir un ou plusieurs pisteurs que nous n'avons pas encore identifiés. Si vous avez des doutes au sujet de ce rapport, contactez-nous à contact@exodus-privacy.eu.org.

13 Permissions

Nous avons trouvé les permissions suivantes dans cette application :

	ACCESS_COARSE_LOCATION (android.permission) *Access Approximate Location (Network-based)*	Dangerous
	ACCESS_FINE_LOCATION (android.permission) *Access Precise Location (GPS And Network-based)*	Dangerous
	ACCESS_NETWORK_STATE (android.permission) *View Network Connections*	Normal
	ACCESS_WIFI_STATE (android.permission) *View Wi-Fi Connections*	Normal
	GET_ACCOUNTS (android.permission) *Find Accounts On The Device*	Dangerous
	INTERNET (android.permission) *Have Full Network Access*	Normal
	RECEIVE_BOOT_COMPLETED (android.permission) *Run At Startup*	Normal
	VIBRATE (android.permission) *Control Vibration*	Normal
	WAKE_LOCK (android.permission) *Prevent Phone From Sleeping*	Normal
	WRITE_EXTERNAL_STORAGE (android.permission) *Modify Or Delete The Contents Of Your SD Card*	Dangerous
	SET_ALARM (com.android.alarm.permission) *Set An Alarm*	Normal
	RECEIVE (com.google.android.c2dm.permission)	Unknown
	C2D_MESSAGE (fr.cityway.android.creaastuce.permission)	Unknown

RGPD

40 ans

- **1978** loi Informatique et Liberté ⟶ Cnil
- **1995** directive 95/46/CE
- **2009** Paquet Télécom
- **2016** loi pour une République Numérique
- **2018** entrée en vigueur du RGPD le 25 mai

RGPD

- **R**ègleme**n**t **G**énéral de **P**rotection des **D**onnées
- Règlement européen relatif à
  - la protection des personnes physiques
  - à l'égard du **traitement** des **données à caractères personnel**
  - et à leur libre circulation

Traitement

- Toute opération
  collecte, enregistrement, organisation, structuration, conservation, adaptation, modification, extraction, consultation, utilisation, communication, diffusion, mise à disposition, rapprochement, interconnexion, limitation, effacement, destruction
- Manuelle ou automatisée

Données à caractère personnel

- Toute information...
  quel que soit le support (électronique, papier...)
- ... se rapportant à une personne physique identifiée ou identifiable
- **Dès qu'une quantité d'informations permet d'identifier une personne, ce sont des données à caractère personnel!**

Principes

- Consentement
  obligatoire, explicite, facile à comprendre
- Finalités déterminés
  précises, explicites, légitimes
- Durée de conservation
  limitée, raisonnable
- Minimisation
  seules les données nécessaires au traitement sont collectées
- Protection
  par défaut et par design, technique et organisationnelle

Droits et obligations

Droits des personnes concernées

- Droit à l'information
- Droit d'accès
- Droit de rectification
- Droit à l'effacement
- Droit à la limitation du traitement
- Droit à la portabilité
- Droit d'opposition

Obligations du responsable du traitement

- Délégué à la Protection des Données (DPO)
  successeur du correspondant Informatique et Liberté
- Preuve que les traitements respectent le RGPD
- Maintenance du registre
  description des traitements, des finalités, des données conservées, de la durée de conservation...
- Étude d'impact
  en cas de risque important
- Certification et code de conduite

Collecte de données à caractère personnel

Concrètement

- Le responsable vous informe du traitement dont vous allez être l'objet
- Il doit fournir les informations
  - d'une façon concise,
  - transparente,
  - compréhensible,
  - aisément accessible,
  - en des termes clairs et simples

Informations à fournir 1/2

- Identité, coordonnées du responsable
- Finalités et base juridique du traitement
  le cas échéant, les intérêts légitimes poursuivis
- Destinataires
- Durée de conservation
- Droit d’accès à vos données
  ainsi que la rectification, l’effacement ou la limitation du traitement

Informations à fournir 2/2

- Droit de retirer son consentement
- Droit d'introduire une réclamation
- Caractère de la collecte
  réglementaire ou contractuel ? Conditionne-t-il la conclusion d'un contrat ? Quelles sont les conséquences en cas de non-fourniture des données ?
- Existence d'une prise de décision automatisée
  Y compris un profilage (ex. : Google Analytics)
- Les autres finalités

Merci d’avoir tenu jusque là!

Quelques liens 1

- Données collectées par les apps – Exodus
  https://reports.exodus-privacy.eu.org/reports/apps/
- Dégooglisons Internet – Framasoftt
  https://degooglisons-internet.org/
- Voir l'empreinte de son navigateur – Am I unique
  https://amiunique.org/
- Qui surveille votre navigation ? – Greffon Ghostery
  https://addons.mozilla.org/fr/firefox/addon/ghostery/

Quelques liens 2

- Rien à cacher ? – Nothing to hide
  https://nothingtohidedoc.wordpress.com/nothing-to-hide-french-high-quality/
- Faire une demande Cada – Xavier Berne
  https://www.nextinpact.com/news/105587-comment-faire-demande-cada-dacces-aux-documents-administratifs.htm
- Google utilise un mouchard pour géolocaliser les utilisateurs Android – Anaïs Cherif
  http://www.latribune.fr/technos-medias/google-utilise-un-mouchard-pour-geolocaliser-les-utilisateurs-android-759095.html
- Le RGPD – EUR-Lex
  http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679

Quelques liens 3

- Pourquoi vous ne devriez pas utiliser Google Chrome
  http://www.revoltenumerique.herbesfolles.org/2014/04/23/pourquoi-vous-ne-devriez-pas-utiliser-google-chrome/
- 5 tests faciles pour voir tout ce que Google sait sur vous
  http://www.archimag.com/vie-numerique/2015/11/10/5-tests-faciles-verifier-google-sait-vous-donnees-personnelles
- Liste des 178 acquisitions de Google – Olivier Duffez
  http://www.webrankinfo.com/dossiers/google/acquisitions
- Chiffres-clés de Google 2017 – Olivier Duffez
  http://www.webrankinfo.com/dossiers/google/chiffres-cles

Quelques liens 4

- Moteur de recherche Qwant
  https://www.qwant.com/
- Moteur de recherche DuckDuckGo
  https://duckduckgo.com/