

Comment la recherche vocale et conversationnelle transforme l'algorithme de Google...

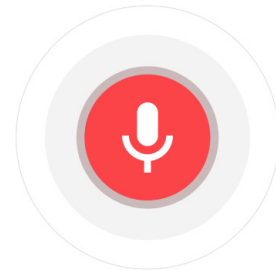


Par Philippe Yonnet

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Nous avons certainement tous testé une fois au moins la recherche vocale de Google, qui devient petit à petit une recherche conversationnelle, cherchant à fournir une réponse la plus pertinente possible au travers d'un aller-retour de questions-réponses avec l'utilisateur. Cette façon d'interagir avec l'être humain par la voix a été initiée en 2008 par Google et est maintenant au cœur de ses algorithmes, Hummingbird et RankBrain notamment. De ce fait, elle influe obligatoirement sur la façon dont nous devons concevoir notre métier de SEO à l'avenir. Décryptage...

Intégrer avec son ordinateur en lui posant des questions en langage naturel comme dans les épisodes de Star Trek est un vieux rêve que les ingénieurs cherchent maintenant à faire entrer dans la réalité. Chez Google, ce projet a pris la forme de la fonctionnalité "Google Voice Search", la recherche vocale de Google.



Google Voice Search est apparu comme une nouvelle fonctionnalité sur l'application mobile Google dès 2008. Disponible d'abord en anglais uniquement, Google l'a ensuite fait évoluer pour qu'elle fonctionne dans des dizaines d'autres langues (le support du français date de 2010). Fin 2011, une version fortement améliorée de cette fonctionnalité est intégrée dans Android Jelly Bean, et fusionnée avec le projet d'assistant personnel Google Now. La filiation avec Star Trek est soulignée par le choix du nom de code du projet : Majel, le prénom de l'actrice qui prêtait sa voix à l'ordinateur dans les aventures du Capitaine Kirk.

Depuis lors, Google Voice Search a été étendue sur desktop, intégrée à Chrome, et a pris sa place dans l'expérience de recherche de nombreux internautes. Mais pour permettre que cet outil fonctionne correctement, Google a dû profondément faire évoluer l'architecture de son moteur, le fonctionnement de son requêteur, et même son algorithme de classement. Insidieusement, la recherche vocale est en train de changer considérablement le fonctionnement interne de Google, et, nous allons le voir, la façon de faire du référencement.

Comprendre les questions en langage naturel est un vrai défi

Dans les premières années de fonctionnement de Google Voice Search, son objectif était assez réduit : permettre aux internautes de « dicter » leur requête à la voix. Il s'agissait donc de « reconnaître » les mots clés prononcés par l'utilisateur, et ensuite la requête était traitée de manière traditionnelle dans le moteur.

Mais cette approche « basique » de la recherche vocale a demandé qu'un grand nombre d'obstacles soient éliminés ou contournés, obstacles qui ont empêché une large diffusion de ces technologies dans le passé. De quels écueils parle-t-on ?

- Isoler le son de la voix d'une personne du bruit ambiant est possible, mais pas sans dégradation du taux de reconnaissance. Evidemment, pour une recherche vocale effectuée sur un smartphone, la présence d'un bruit ambiant fort peut être fréquente.
- Créer un système de reconnaissance du langage universel s'est longtemps avéré un défi impossible à relever totalement.
 - o Créer des systèmes fonctionnant dans de nombreuses langues exige des ressources et des investissements importants. Le principal challenge étant en plus de permettre la reconnaissance de deux langues mélangées dans les propos de l'utilisateur (un français utilisant des termes anglais dans sa phrase !). Ce dernier point fait toujours partie des sujets à améliorer sur les systèmes actuels.
 - o Les différences de prononciation entre locuteurs de la même langue créent de nombreuses difficultés : les accents, la prosodie, les différences de timbre, les voix enrouées etc. créent des problèmes complexes à résoudre.
 - o Pour une recherche vocale, il faut reconnaître bien plus que les mots du dictionnaire et leurs formes fléchies (mots avec la marque du genre, du cas, du nombre, verbes conjugués etc.). Il faut aussi, et c'est essentiel pour que les résultats de recherche soient pertinents, reconnaître la prononciation d'entités nommées : noms de personnes, de lieux, de société etc. Ce qui fait exploser la volumétrie de données à exploiter pour la reconnaissance, car le nombre des entités nommées est important, et leur prononciation est souvent variée parmi les locuteurs.
- Techniquement, la quantité de données à embarquer a longtemps été un obstacle pour de nombreuses applications. La reconnaissance de la parole s'appuie depuis des dizaines d'années sur l'emploi combiné de « modèles de langage » (donnant la probabilité pour que le mot B suive le mot A dans une phrase), de « modèles de prononciation » (donnant la probabilité pour qu'un mot soit prononcé d'une certaine façon), et de « modèles acoustiques » (donnant la probabilité pour que les sons analysés correspondent à une prononciation d'une syllabe donnée). Ces modèles s'appuient sur des bases de données dont la taille a longtemps été un obstacle. Au début, la plupart des applications de la reconnaissance automatique de la parole sur smartphone passaient donc par un traitement à distance, et non en local, total ou partiel de la reconnaissance.

Google Voice Search, Siri, Cortana et tous leurs concurrents ont bénéficié des progrès de l'état de l'art à partir des années 2000 dans ce domaine. Google et Microsoft ont également massivement investi dans la recherche sur ces sujets depuis 15 ans, et ont fait beaucoup progresser les modèles et les algorithmes. Le résultat est que le taux d'erreur dans la reconnaissance de la parole de ces systèmes est tombé sous les 10% en conditions de laboratoire, rendant leur utilisation quotidienne possible !

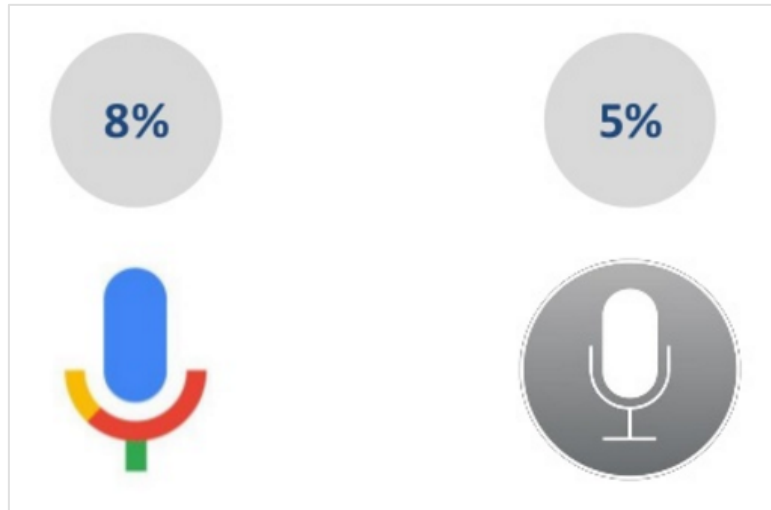


Fig.1. Les taux d'erreurs communiqués par Google pour Google Now (Google I/O mars 2015) et par Apple pour Siri en juin 2015.

L'étape suivante : comprendre le langage naturel

Pour offrir une meilleure expérience de recherche, l'étape suivante est logiquement de permettre à l'utilisateur de formuler ses demandes au moteur sous forme de questions posées en langage naturel. Reconnaître les mots prononcés n'est donc plus suffisant, il faut s'appuyer sur la syntaxe de la phrase pour identifier le sens de la question posée, et ceci, si possible, en temps réel.

Lors du lancement de Google Voice Search en 2008, des outils d'analyse syntaxique robustes et rapides existaient déjà pour l'anglais et de nombreuses langues européennes. Mais pas pour toutes les langues que Google voulait couvrir. La mise au point des différentes versions linguistiques de Google Voice Search (et de Cortana pour Microsoft), a nécessité un gros travail effectué à marche forcée par des équipes de linguistes chez Google et de Microsoft, mais aussi avec l'aide de sociétés partenaires spécialisées.

De la recherche vocale à la recherche conversationnelle

Le « modèle » d'expérience de recherche que Google voulait introduire dans Android Jelly Bean était clairement... l'univers de Star Trek (d'où le nom de code « Majel »).

Amit Singhal, l'ex Vice-Président de recherche et père de l'algorithme de classement de 2001 à 2015, et fan de Star Trek décrivait clairement sa conception d'un téléphone sous Android comme un mélange de « Tricordeur et de Communicateur ». Le projet « Majel » demandait donc que Google Now (et son composant Google Voice Search) permette d'avoir une « conversation » avec le moteur de recherche, pour obtenir des informations, sur le modèle suivant :

- « - Capitaine Kirk : Ordinateur, est-ce que les caractéristiques de la planète Alpha Prime sont compatibles avec la présence de signe de vie ?
- Ordinateur : Alpha Prime est une planète de classe C, et peut potentiellement héberger de la vie. Voulez-vous que je scanne la planète pour détecter des formes de vie ?
- Oui ;
- Je détecte des formes de vie sur le continent au niveau de l'équateur ;
- Ordinateur, l'atmosphère de la planète est-elle respirable pour des humains ?
- L'atmosphère d'Alpha Prime est composée de 17% d'oxygène, 80% d'azote et de traces d'autres gaz neutres. Elle est parfaitement respirable par des humains ;
- Ordinateur, calcule une trajectoire d'approche [avec Alpha Prime comme destination]. Scotty, en vitesse d'impulsion, moteurs au maximum. »

Obtenir des informations en tenant une conversation composée d'échanges réciproques de questions réponses, correspond au concept de « recherche conversationnelle ». Google a cherché dès 2010 à ajouter des fonctionnalités sophistiquées de recherche conversationnelle dans Google Voice Search. Et cela a abouti à l'implémentation actuelle.



Fig.2. Si vous posez la question : « Ok Google, quel est l'âge de Barack Obama ? », Google comprend le sens de la question, et fournit directement la réponse. Ici, sont à l'œuvre : la reconnaissance de la parole, et une analyse syntaxique pour identifier que la réponse à la requête est un attribut de l'entité « Barack Obama ».

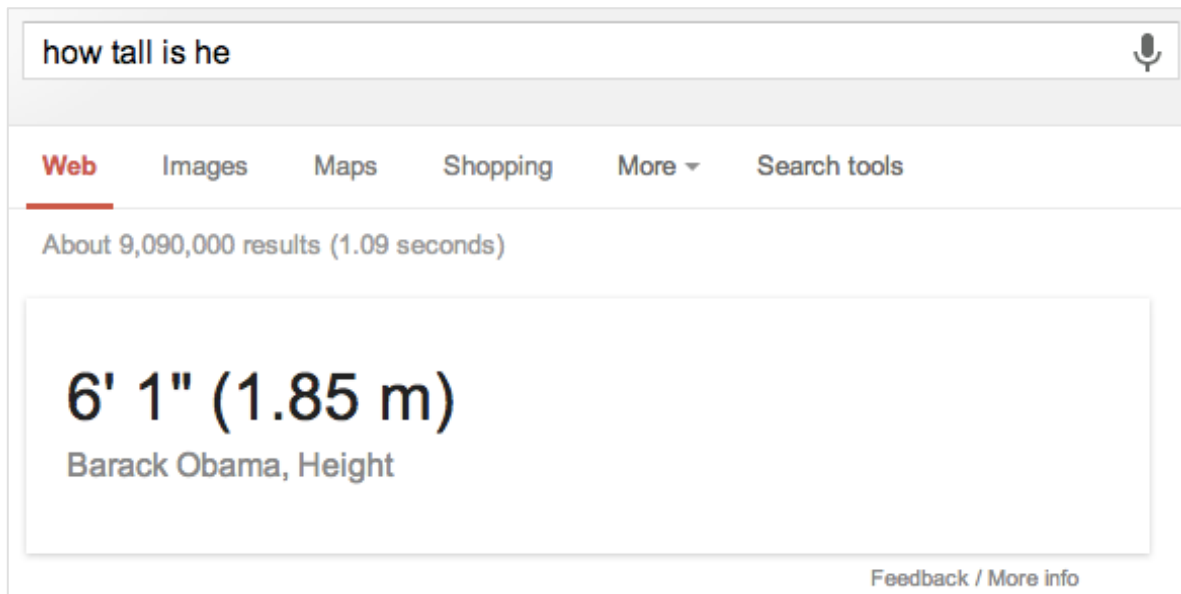


Fig.3. Si vous continuez à poser des questions de manière naturelle, la prochaine question devient « quelle est sa taille ? ». Le système tient compte du contexte dans lequel est posée la question, et identifie que "sa" fait référence à Barack Obama. Cet exemple a été présenté en 2013 par Google lors d'un Google I/O.

Ce mode de recherche demande de stocker l'historique de la conversation et d'analyser le contenu pour réaliser ce que les linguistes appellent « résolution des coréférences » ou « résolution des anaphores ». Dans l'exemple tiré de Star Trek, la planète est une coréférence, c'est-à-dire une autre appellation pour l'entité Alpha Prime (comme Barack Obama peut aussi être appelé POTUS ou le Président des Etats- Unis). Et « he » ou (« sa » en français) dans l'illustration ci-dessus sont des anaphores, les pronoms font référence à une entité désignée explicitement auparavant.

La recherche conversationnelle demande aussi d'exploiter des informations de contexte, car certaines questions peuvent contenir des références à des entités qui n'ont pas été explicitement désignées, mais qui sont connues implicitement (comme l'endroit où est placé l'internaute, information de contexte apportée par la géolocalisation du smartphone).

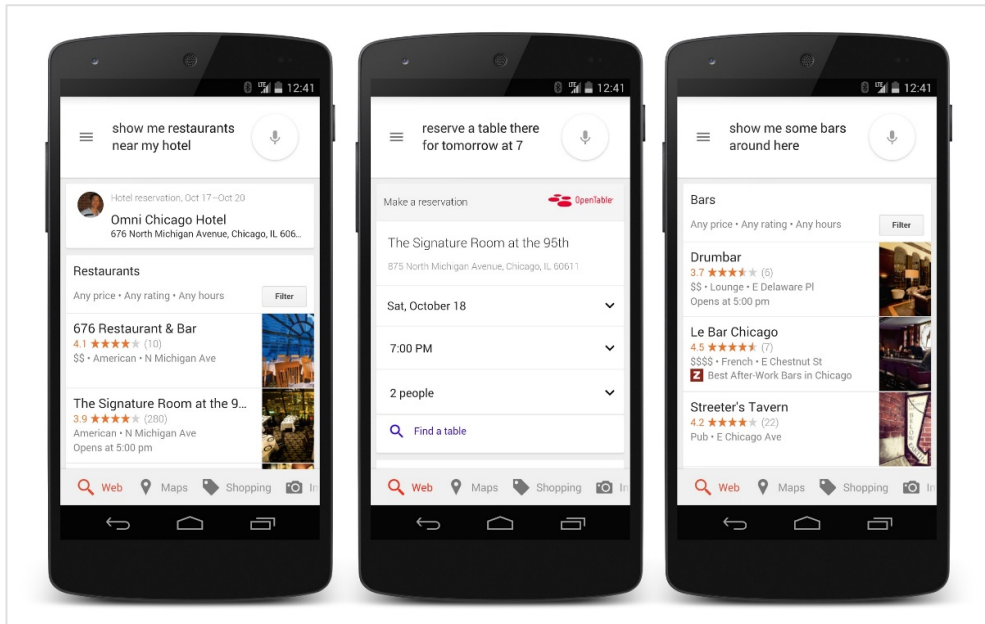


Fig.4. Résolution des anaphores sur des entités implicites avec Google : « affiche les restaurants à proximité de mon hôtel ». « Réserve une table ici pour demain à 7h... »

La recherche conversationnelle exige un "moteur de réponses"

L'expérience de recherche à la Star Trek exige par ailleurs que le système fournisse des réponses à partir d'une base de connaissances. Dans Google Now, l'interface pour le retour d'informations est à la fois visuelle (les « cartes ») et vocale. Les cartes présentent des faits, des chiffres, des informations, sous formes synthétiques. L'interface vocale fournit la même information, sous une forme encore plus abrégée. Ces interfaces sont pratiques, parfaitement adaptée pour un usage « mains libres » sur un smartphone ou pour l'internet des objets. Mais ces interfaces limitent la quantité d'information qui peut être délivrée en réponse à la requête. L'idéal est donc que le système fournisse LA réponse, et non DES réponses.

La recherche conversationnelle fait donc disparaître les « dix liens bleus », soit totalement, soit la relègue complètement au second plan. Les « dix liens bleus » n'apparaissent plus que dans les cas où le système :

- Ne sait pas renvoyer de réponses ;
- Ou ne comprend pas bien le sens de la question.

La première transformation de l'algorithme de Google : Hummingbird et l'indexation des concepts

Pour le moment, soit début 2016, une majorité des requêtes vocales font encore apparaître une page de résultats « normale » de Google. Mais lorsque la question est posée en langage naturel, un nouveau challenge apparait : comment continuer à fournir

des résultats pertinents quand l'internaute utilise cette voie (voix ?) pour poser ses questions ?

Les formulations en langage naturel posent trois problèmes :

- Elles sont le plus souvent plus ambiguës (à cause des références implicites), même si elles peuvent aussi être parfois plus précises qu'une saisie de mots clés mal choisis.
- Elles sont extrêmement variées dans leur tournure. Cette caractéristique rend l'analyse des sessions de requête inopérante pour détecter quel type de page ou de résultats renvoyer vers certaines requêtes tapées en langage naturel.
- Une question en langage naturel n'appelle plus un « mot clé » ou une « expression clé ». Mais plutôt une information ou une action.

Cette dernière notion, Google la regroupe sous le terme générique de « concept ». Une question en langage naturel appelle donc un « concept » comme réponse. Et pour retourner les pages associées au « concept » appelé par la question, cela signifie qu'il faut savoir associer des concepts à des documents. La recherche conversationnelle a donc conduit Google à revoir complètement son architecture en 2013, pour permettre :

- Le stockage des relations termes <-> concepts, et documents <-> concepts ;
- Et le traitement des requêtes en exploitant les concepts.

Cette nouvelle architecture est connue sous le nom de Hummingbird, elle a été présentée lors de l'évènement organisé pour l'anniversaire de Google en Septembre 2013.

Depuis Hummingbird, lorsqu'une requête est adressée au moteur, le système fait une « expansion de requête » (il cherche à retourner des informations complémentaires à l'aide de requêtes complémentaires) en cherchant des documents associés au concept appelé par la requête. Hummingbird permet donc la remontée de pages qui sont (théoriquement) une bonne réponse à la question posée, sans que ces documents soient vraiment similaires, au niveau lexical, avec la requête.

La seconde transformation de l'algo : Rankbrain et la sémantique au coeur de l'algo

Les progrès de Google en Intelligence Artificielle ont fait progresser la recherche vocale de manière spectaculaire depuis 2013. L'approche en « word embedding » (décrite dans notre article du mois dernier) a permis d'améliorer les modèles de langue de façon spectaculaire. L'emploi de nouvelles générations de systèmes à base de réseaux de neurones, et de *machine learning* ont contribué à améliorer les modèles acoustiques et les modèles de prononciation.

Rankbrain, implémenté depuis l'année dernière, et qui exploite lui aussi les « word embeddings » et des modèles de langues améliorés, apporte aussi une amélioration spectaculaire aux performances de l'algorithme de classement « post » Hummingbird. Rankbrain est particulièrement performant dans l'identification des « vraies » relations de proximité sémantique entre les termes. Cela signifie que Rankbrain et Word2vec apportent un vrai gain qualitatif dans l'identification des pages pertinentes, sans

s'appuyer sur des occurrences exactes des mots clés de la requête. La proximité sémantique prend donc le pas sur la similarité lexicale.

Combien d'internautes utilisent vraiment aujourd'hui la recherche vocale ?

Google, Apple et Microsoft ont communiqué peu de chiffres précis sur le niveau d'adoption de la recherche vocale par les internautes. Mais quelques études indépendantes, et les quelques informations lâchées par les GAFAs avec parcimonie montrent que l'adoption de la recherche vocale par les internautes est en progrès spectaculaire.

Selon une étude de MakeMeld, 33% des internautes US utilisent la recherche vocale au moins une fois par semaine et 22% chaque jour. L'utilisation quotidienne a progressé de 4% en 3 mois !

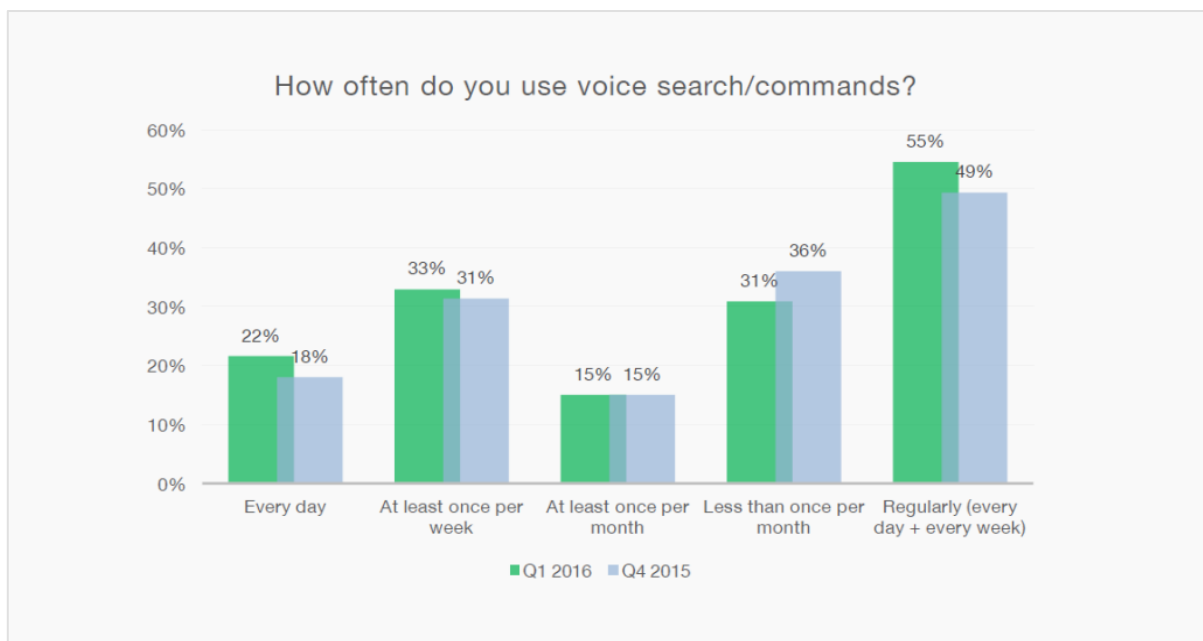


Fig. 5. Etude Mindmeld – Q1 2016 : quel est la fréquence d'utilisation de la recherche vocale par les internautes ?

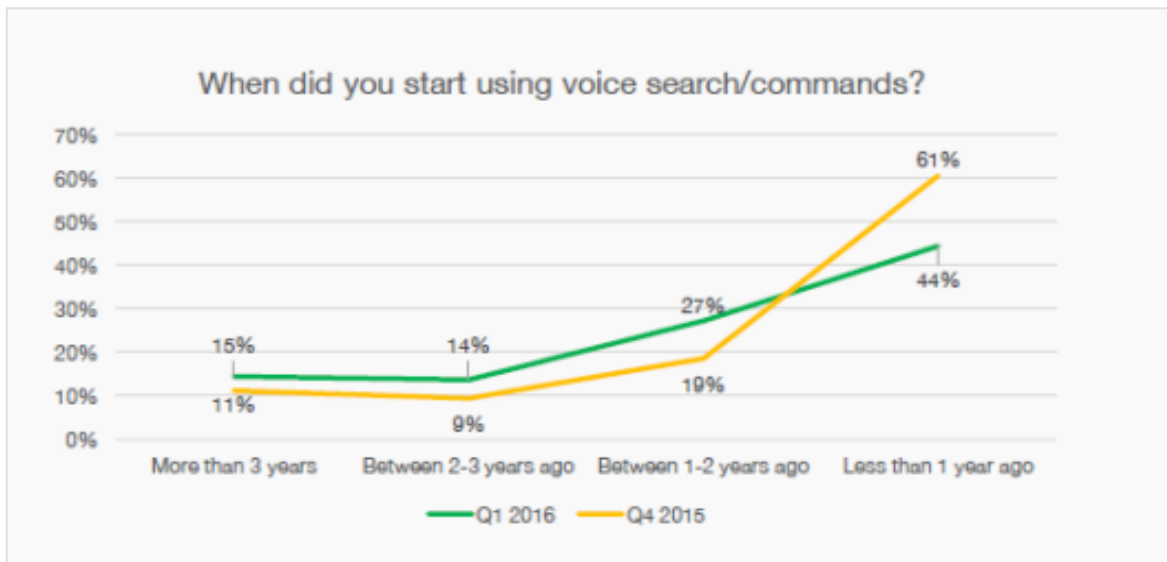


Fig.6. Etude MindmeID – Q1 2016 : quand les internautes ont-ils commencé à utiliser la recherche vocale?

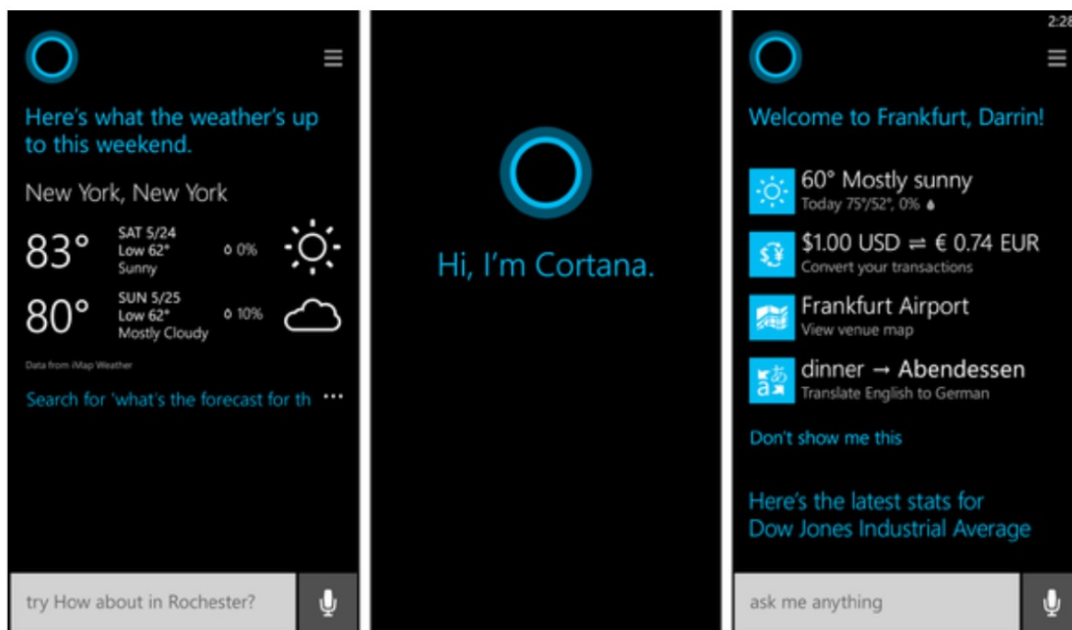


Fig. 7. La présence par défaut de Cortana dans Windows 10 a dopé la recherche vocale sur Bing : 25% des requêtes sur Bing sont des recherches vocales !

Le taux d'adoption chez les teenagers est encore plus spectaculaire. Une étude de Google réalisée en oct. 2014 montre que cette catégorie d'utilisateurs ont rapidement adopté ce mode de recherche (55% l'utilisent quotidiennement), notamment à cause des usages multi-écrans et mains libres.

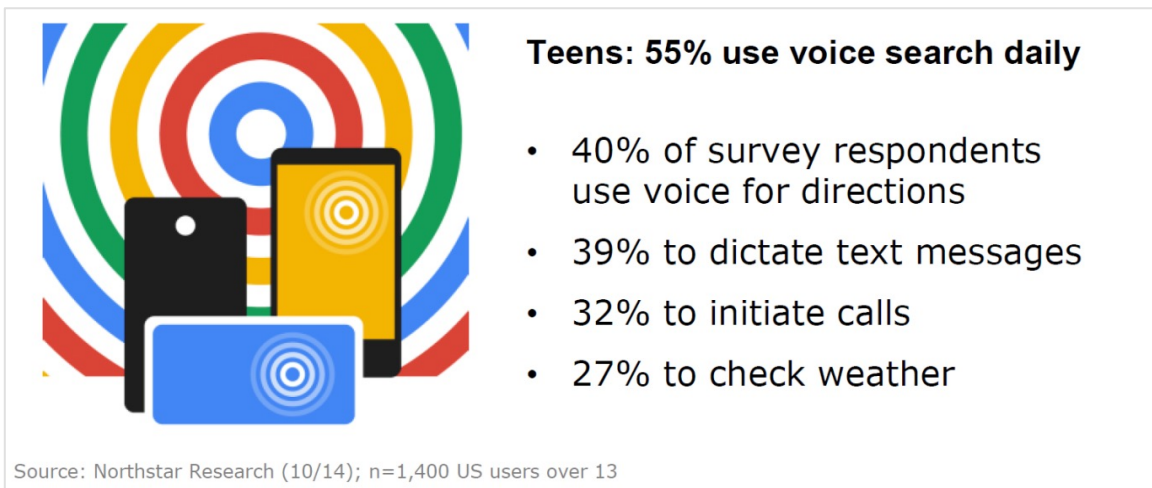


Fig.8. Utilisation de la recherche vocale chez les jeunes, étude Google (2014).

Qu'est-ce que cela va changer dans les méthodes de référencement ?

Il n'y a toujours qu'un seul algorithme de classement qui gère les réponses fournies par le moteur, mais plusieurs mécanismes sont à l'œuvre en parallèle en fonction d'une part de la détection de l'intention de l'utilisateur qui a formulé la requête, mais aussi, de la façon dont la requête a été formulée (frappe de mots clés ou questions en langage naturel, au clavier ou par recherche vocale/conversationnelle) et où elle a été tapée (sur un smartphone, avec un accès élargi aux informations de contexte et aux informations personnelles, ou sur un desktop, avec un accès limité au contexte ?).

Le fonctionnement du processeur de requêtes donne la part belle à différentes parties de l'algorithme et de l'architecture en fonction du mode de formulation : une recherche vocale, avec une question en langage naturel, va déclencher une expansion de requête exploitant les infos stockées dans l'architecture Hummingbird, en exploitant également les possibilités de détection de termes sémantiquement similaires fournis par Rankbrain.

La conclusion à en tirer est simple : la recherche conversationnelle produit de plus en plus de cas où :

- Soit la réponse est donnée directement par Google. Dans ce cas on ne peut plus capter le trafic créé par ces recherches !
- Soit la réponse fournie par le « moteur de réponses » est un « featured snippet » (voir notre article de mars 2016 dans cette lettre à leur sujet). Dans ce cas, on peut espérer capter ce trafic, mais malheur à tous les autres qui n'ont pas été retenus comme source pour la réponse.
- Soit Google affiche des liens bleus, mais dans ce type de recherches, de plus en plus de documents seront choisis en raison de leur proximité sémantique avec le concept appelé par la question.

Clairement, la recherche conversationnelle contribue à accélérer la mutation de l'algorithme de Google vers une expérience de recherche plus « naturelle », dans laquelle la pertinence des résultats est améliorée par des fonctionnalités sémantiques plus avancées. En clair, le moteur « comprend » mieux le contenu des pages, et cela veut dire que des contenus artificiellement optimisés sont moins efficaces face à des outils plus « sémantiques ».

Les méthodes d'optimisation « on page » traditionnelles vont voir de plus en plus leur efficacité s'éroder. Aujourd'hui, elles fonctionnent encore. Mais demain, il vaudra mieux s'attacher avant tout à créer une bonne page répondant correctement aux requêtes fréquentes des internautes, que de faire du keyword stuffing ou de l'optimisation de balises <title>...

L'avenir du référencement ressemblera-t-il à de l'optimisation pour l'interface de l'ordinateur de Star Trek ? Ce qui est sûr, c'est qu'il faut commencer à aller « *boldly, where no SEO has gone before* ».

Bibliographie

Recherche vocale par Google :

<http://googleresearch.blogspot.fr/2009/11/google-search-by-voice-learns-mandarin.html>

Sur Hummingbird :

<http://insights.wired.com/profiles/blogs/google-hummingbird-where-no-search-has-gone-before#axzz48E1OZd74>

<http://searchengineland.com/google-hummingbird-172816>

Google Search by Voice: A case study

<http://static.googleusercontent.com/media/research.google.com/fr//pubs/archive/36340.pdf>

L'article Abondance : Sur Bing, un quart des recherches sont vocales !

<http://www.abondance.com/actualites/20160510-16567-sur-bing-un-quart-des-recherches-sont-vocales.html>

La conférence de Yann Sauvageon (Synodiance, groupe Search Foresight) au SEO Campus de Paris :

http://fr.slideshare.net/Synodiance/synodiance-recherche-vocale-seo-campus-paris-07042016?qid=215b606b-6649-4c9f-8083-5647784da838&v=&b=&from_search=1

Voir aussi :

Lettre Recherche et Référencement (novembre 2014) : Cortana Siri et Google Now vont-ils changer l'univers de la recherche d'information ? :

<http://recherche-referencement.abondance.com/2014/11/cortana-siri-et-google-now-vont-ils.html>

Lettre Recherche et Référencement (mai et juin 2015) : Speech Processing et SEO : l'avenir du référencement audio (1ère et 2ème partie) :

<http://recherche-referencement.abondance.com/2015/05/speech-processing-et-seo-lavenir-du.html>

<http://recherche-referencement.abondance.com/2015/06/speech-processing-et-seo-lavenir-du.html>



Philippe YONNET, *Directeur Général de l'agence Search-Foresight, groupe My Media*
(<http://www.search-foresight.com>)