

# Google Now i Siri

---

## Assistents Personals Virtuals

**Iñaki Frisón Pujol**  
**Adrian Hintze Molero**  
**Marti Zamora Casals**

**18/12/2012**

# Índex

1. Introducció als assistents personals.....	2
1.1. Precedents.....	2
2. Siri i Google Now .....	3
2.1. Siri.....	3
2.1.1 Orígens .....	3
2.1.2 Estat actual .....	4
2.2. Google Now.....	4
2.2.1 Orígens .....	4
2.2.2 Estat actual.....	5
3. Tecnologies que utilitzen .....	5
3.1. Reconeixement de veu.....	5
3.1.1. Historia .....	5
3.1.2. Mètode.....	6
3.2. Processament del llenguatge natural.....	7
3.2.1. Mètodes al llarg de la historia.....	7
3.2.2. Question Answering .....	8
3.3. Diferències i detalls i implementació de les tecnologies de cada un.....	8
3.3.1 Siri:.....	8
3.3.2 Google Now.....	10
3.4 Limitacions de la tecnologia .....	11
4 Impacte.....	12
4.1. Perquè Google i Apple estan interessats en tenir els seus propis i no deixen que siguin aplicacions externes.....	12
4.2. Acceptació .....	12
4.3. Que aporten aquests productes .....	12
4.3.1 per la empresa (riscos i beneficis).....	13
4.3.2 Per a la gent que l'utilitza (riscos i beneficis) .....	14
5.Conclusions .....	14
6.Bibliografia .....	15

## **1. Introducció als assistents personals**

Un assistent personal, en tecnologia, és un programa que emula el comportament d'un assistent personal real, és a dir, intenta fer recomanacions i donar informació que pugui resultar útil a l'usuari. Per poder portar a terme aquesta tasca es nodreix de grans quantitats d'informació (localització, memòria a llarg, mitjà i curt termini, preferències, notícies, temps...) que han de ser processades adequadament mitjançant tècniques d'intel·ligència artificial.

En el darrer any les grans empreses punteres en el sector tecnològic s'han bolcat en el desenvolupament d'assistents personals que són incorporats als seus respectius sistemes operatius mòbils. Això és degut a una evolució lògica en l'intent de tractar i fer més accessibles les ingents quantitats de dades que són emmagatzemades en els seus servidors des de fa temps. Fins ara no s'havia pogut portar a terme per dues raons primordials:

La falta de potència de càlcul dels ordinadors en general i el fet que les tècniques d'intel·ligència artificial necessàries encara no estaven suficientment desenvolupades.

En aquest document s'expliquen i comparen dos productes tecnològics que fan la funció d'assistent personal Siri i Google Now. Desenvolupats per Apple i Google respectivament. Hem pensat que era interessant estudiar els dos i comparar-los ja que són productes molt semblants i en el fons tenen un objectiu molt similar tant des del punt de vista tecnològic com des del punt de vista empresarial ja que les dues empreses són les impulsores dels dos sistemes operatius mòbils més utilitzats actualment.

Un altre punt important que s'analitza són les diverses tècniques d'intel·ligència artificial en que s'han basat i quines tècniques i "trucs" han utilitzat per implementar-les.

El darrer punt analitzat ha estat l'impacte que han causat i que encara causen aquests productes des de els motius que han portat a aquestes empreses per crear una aplicació d'aquest tipus (en gairebé el mateix període de temps) fins a els beneficis que aprofiten tant des del punt de vista empresarial com des del punt de vista d'usuari. També s'analitzen els riscos que comporta l'ús d'aquestes noves tecnologies i la acumulació de dades de tot tipus per part de les empreses dels seus clients.

### **1.1. Precedents**

Ja abans que apareguessin Siri i Google Now, hi ha hagut intents de fer "assistents personals" amb una finalitat semblant tot i que més especialitzats, i altres productes que fan servir tècniques d'intel·ligència artificial utilitzades en els assistents moderns.

Tots aquests productes (comercials o no) han estat la base per als productes que estem veient aparèixer actualment, com poden Siri, Google Now, S Voice.

Per exemple tenim el cas del "Clip" de Microsoft Office. Aquest assistent intel·ligent ja intentava donar consells als usuaris a finals de 1997. Malgrat tot mai va tenir gaire èxit i el seu ús no es va estendre entre els usuaris. Aquest fet és degut a que Microsoft no s'havia plantejat prou bé el comportament "social" que havia de tenir l'assistent, tal com va dir el professor

Clifford Nass (un expert en interacció persona-màquina). Finalment el producte va ser retirat l'any 2007 i la nova versió de Microsoft Office (la versió 2007) ja no incloïa aquesta funció.

Un altre sistema anterior és el reconeixedor de veu del sistema operatiu Windows (que està integrat des de la versió Windows Vista) aquest sistema té com a objectiu eliminar al màxim possible l'ús del teclat i el ratolí permetent a l'usuari comunicar-se amb el PC mitjançant la veu (per tant, de manera similar als assistents personals, necessita ser capaç de reconèixer la veu humana). La diferència principal amb els assistents moderns es que és un sistema molt més limitat. Ja que no té la mateixa capacitat d'entendre llenguatge natural i necessita una gran quantitat d'entrenament. A més el seu ús està limitat ja que només està preparat per interactuar amb un número fix i limitat d'aplicacions (per a la resta fa servir comandaments genèriques). I no té la capacitat de respondre a consultes que tenen productes com Siri o Google Now.

Un altre tipus d'aplicacions predecessores als assistents moderns són els programes de xat amb la computadora, aquests programes (els primers dels quals van ser desenvolupats durant els anys 70) intenten entendre el llenguatge natural de les persones per tal de donar respostes coherents en aquest llenguatge i intentar simular una conversació, de manera semblant

a com el assistents moderns necessiten entendre preguntes i altres ordres de l'usuari donades en llenguatge natural i donar respostes coherents.

## 2. Siri i Google Now

### 2.1. Siri



Siri és l'assistent personal integrat a l'iOS d'Apple des de la versió iOS 5 per al seu producte iPhone i des de la versió iOS 6 per a iPad. Es caracteritza per fer servir una interfície de llenguatge natural per tal de portar a cap nombroses accions per l'usuari, com contestar a preguntes, fer recomanacions, etc, mitjançant una sèrie de serveis web. L'aplicació aprèn amb el temps i es va adaptant a les preferències de l'usuari.

#### 2.1.1 Orígens

Siri té els seus orígens en una aplicació per iPhone creada per l'empresa Siri Inc. al 2007 per Dag Kittlaus. L'empresa va ser comprada per Apple al 2010 i el producte va ser integrat per primer cop a l'iPhone 4S presentat a finals de 2011. Siri estava pensat en un inici com un sistema multiplataforma, però la compra per part d'Apple va limitar el seu ús als seus productes.

### 2.1.2 Estat actual

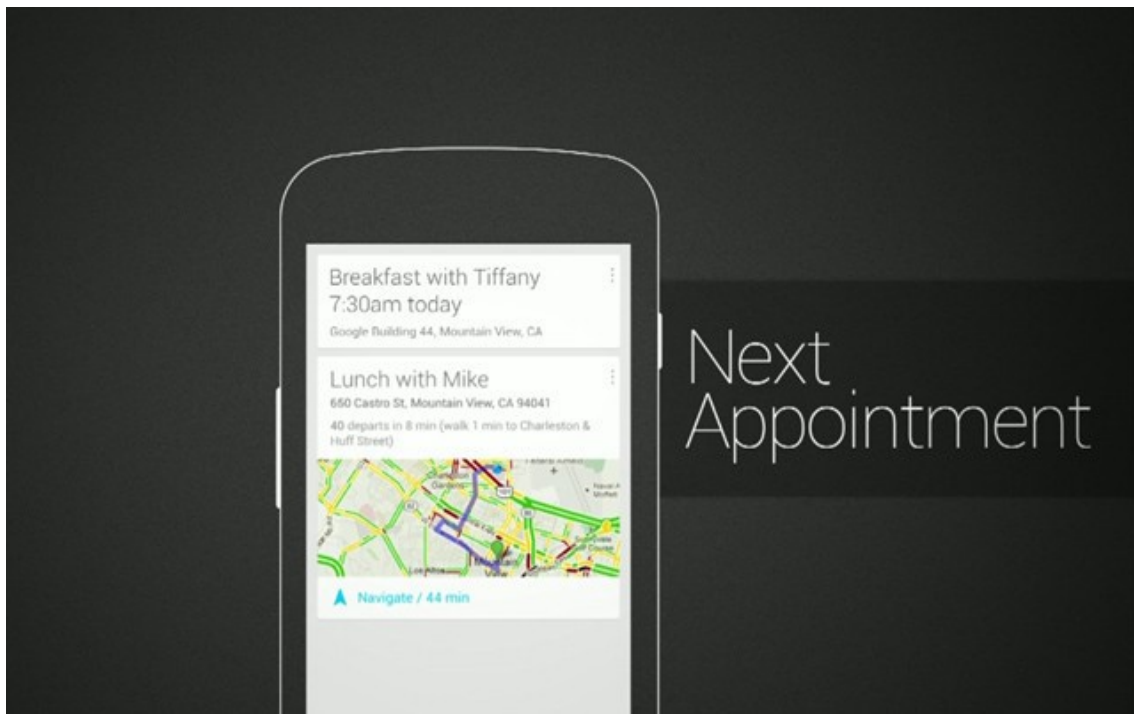
Actualment Siri està integrat a tots els nous dispositius mòbils d'Apple i permet interactuar mitjançant la veu amb la gran majoria de les seves aplicacions (com el seu reproductor de música iTunes, el calendari...) i moltes altres creades per terceres empreses com poden ser Google Maps i Wolfram Alpha, d'entre altres.

Un dels principals problemes de Siri és que actualment només és compatible amb un nombre limitat d'idiomes (el nombre és encara més limitat en la primera versió d'iOS que va integrar Siri), i fins i tot, dins d'un mateix idioma està limitat a uns pocs dialectes. Actualment, entre llengües i dialectes en són suportats 19.

Un altre problema de l'aplicació és que actualment la seva funcionalitat està molt limitada fora dels E.E.U.U. i Canadà, restringint, per exemple, la capacitat de buscar serveis com restaurants o cinemes fora d'aquests països.

## 2.2. Google Now

Google Now és un assistent personal disponible per a dispositius mòbils que utilitzen el sistema operatiu Android (desenvolupat i impulsat per l'empresa Google). Utilitza una interfície de llenguatge natural per contestar a preguntes de l'usuari, fer-li recomanacions, etc, mitjançant el Knowledge Graph de Google. Google Now aprèn de l'usuari i li proporciona informació de forma automàtica en funció dels seus hàbits de cerca (intenta predir quin tipus d'informació l'usuari necessitarà o voldrà en un futur). Google Now va passar a estar disponible per primer cop a la versió 4.1 d'Android (Jelly Bean).



### 2.2.1 Orígens

Google Now ha estat desenvolupat per Google com a extensió de la seva aplicació Google Search per a Android. Per implementar-lo Google va adquirir al 2010 l'empresa Metaweb, una

empresa dedicada a recollir informació sobre el món i crear així una gran base de coneixement (anomenada Freebase) que va incloure al Google Knowledge Graph.

### **2.2.2 Estat actual**

Google Now analitza diferent informació de l'usuari com accions que realitza habitualment (cerques, llocs reals que acostuma a visitar, etc) a més d'informació emmagatzemada als seus dispositius, com pot ser el correu electrònic, calendaris... Seguidament, utilitza tot aquest coneixement per mostrar informació rellevant a l'usuari en forma de "cards", cada "card" és un mòdul diferent de Google Now i s'especialitza en un apartat diferent (per exemple en l'anàlisi d'e-mails), l'aspecte més interessant per l'usuari d'aquest sistema de cards és que la informació mostrada és personal, actual i rellevant i intenta avançar-se a les peticions que farà l'usuari en un futur immediat (estalviant-li així feina i temps). Actualment existeixen 24 mòduls diferents encara que aquest número creix amb freqüència (ja que aquest producte és molt recent i està en ple desenvolupament).

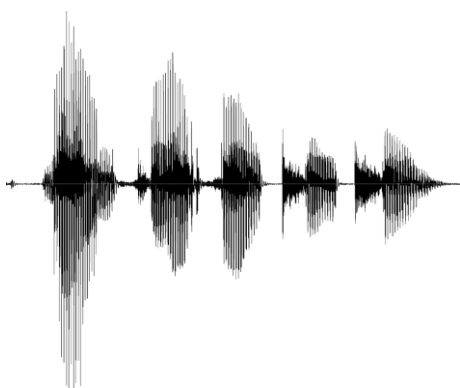
De forma semblant a Siri d'Apple el número de llengües i dialectes que reconeix Google Now és limitat, encara que el seu número és més elevat amb 42 llengües i dialectes diferents.

També pateix de limitacions en alguns països (moltes de les funcionalitats més recents que tenen a veure amb la cerca de serveis com restaurants no són compatibles fora dels Estats Units d'Amèrica).

## **3. Tecnologies que utilitzen**

Les tecnologies principals que utilitzen aquests dos assistents es el reconeixement de veu per a entendre a l'usuari i el processament de llenguatge natural per respondre preguntes.

### **3.1. Reconeixement de veu**



El reconeixement de veu es la tecnologia que forma part de la intel·ligència artificial i té com a objectiu permetre la comunicació parlada entre persones i ordinadors.

Es tracta d'un problema realment difícil computacionalment, però les eines actuals són increïblement sofisticades i tenen uns índex de precisió realment bons.

#### **3.1.1. Historia**

El reconeixement de veu ha estat popularitzat molt recentment per aplicacions com Siri gràcies a que aquets sistemes han arribat a un punt de precisió increïble. Tot i així les investigacions en aquest camp que ens han permès tenir aquets sistemes tant perfectes va començar fa mes de mig segle.

En 1952 els Laboratoris Bell van presentar “Audrey” capaç de reconèixer els dígets de l’1 al 10 dits per una certa persona. Deu anys més tard IBM mostrava al públic “Shoebbox” que podia entendre 16 paraules en anglès. Durant els anys setanta aquets programes de reconeixement de veu van millorar de manera molt important, sent capaçs de reconèixer varis centenars de paraules o interpretar la veu de més d’una persona. Als anys vuitanta els reconeixedors de veu van millorar la seva precisió i van augmentar el nombre de paraules que eren capaçs de interpretar fins a varis milers, gracies a un nou mètode estadístic (els Hidden Markov Models). Els primers sistemes comercial són dels anys noranta (Dragon Dictate) i durant aquesta dècada van anar millorant (Dragon Natural Speaking, etc) però la seva precisió semblava que no podia superar el 80%. Va ser Google que l’any 2008 va revolucionar aquest camp al llençar un aplicació per a mòbils que permetia fer cerques en el seu buscador amb la veu. Google aconseguia millorar notablement la precisió combinant la potencia dels seus data centers, milers d’hores d’entrenament i la informació recol·lectada sobre els bilions de cerques realitzades en el seu buscador.

### **3.1.2. Mètode**

La majoria de gravacions de veu estan mostrejades amb una freqüència de 8kHz, 16kHz (telefonía, Veu sobre IP), 44,1kHz (CDs de música) o 48kHz (pel·lícules, DVDs, etc). La majoria de mòbils actuals tenen micròfons optimitzats per captar veu i són capaçs de gravar en freqüències altes amb bona qualitat. Alguns tenen, a més, micròfons secundaris que permeten aplicar tecnologies de cancel·lació de soroll per obtenir gravacions encara més clares.

#### Divisió en fragments

La quantitat d’informació es massa extensa i la major part és completament irrellevant per al reconeixement de veu. La solució més comuna es dividir la mostra en fragments de la mateixa mida i escollir paràmetres significatius de cada fragment. Estudis sobre la intel·ligibilitat de la parla senyalen que la mida d’aquests fragments ha de ser d’almenys 10 ms. Si els fragments són més grans de 20ms es produeix una distorsió massa alta. En canvi, quant més petits són, més n’hi ha i es necessita més capacitat de càlcul.

#### Extracció de característiques (MFCC)

De cada un d’aquets fragments s’extreuen les característiques que el descriuen. Experimentalment s’ha determinat que es necessari un vector de uns 40 valors. Aquests valors son funcions molt diverses. Un dels conjunts de valors més utilitzats son els MFCC (Mel Frequency Cepstral Coefficients (coeficients cepstrals en las freqüències de Mel)), però n’hi ha d’altres com els LFCC, LPCC, etc. A més aquestes transformacions de les dades ajuda a reduir la influencia de característiques irrellevants i del soroll.

#### Modelatge acústic (GMMs)

Per determinar quin fonema està sent dit en cada fragment (descriu per un vector de 40 valors) s’utilitza models probabilístics basats en un entrenament previ.

La idea es que aquets entrenaments que contenen moltíssims fonemes dits de manera diferent generen un conjunt de punts en un espai 40 dimensional, i que els punts que representen cada un dels fonema estan bastant agrupats. Així doncs podem determinar amb una certa probabilitat si un fragment es un fonema determinat segons com de a prop estigui a un d’aquets núvols de punts.

Per portar a terme aquesta idea s'utilitzen models gaussians. Cada núvol de punts es mapeja amb una sèrie de distribucions gaussianes formant una barreja de models gaussians (Gaussian Model Mixture). Empíricament es determinen quantes són necessàries i s'entrenen iterativament fins que el model de probabilitats s'adapta prou be. Aquests models ens permeten saber amb quina probabilitat un fragment és un cert fonema.

#### Descodificació (HMMs)

Ara coneixem per a cada fragment les probabilitats de ser un o altre fonema i hem de determinar quina paraula formen una sèrie de fragments. Per fer-ho s'utilitzen els models ocults de Markov (Hidden Markov Model) que són una espècie de graf dirigit amb pesos creats a base d'un entrenament previ. Cada node representa un fonema i el pes de cada aresta les probabilitats de que un fonema segueixi a un altre. Així doncs la paraula més probable és aquella que és més probable segons les probabilitats de transicions i les probabilitats dels fonemes.

Aquests mètodes han estat utilitzats des de els anys 70 per el reconeixement de veu. Hi ha altres aproximacions a aquest problema: com discriminar dinàmicament els fonemes i determinar la seva llargada (Dynamic Time Wrapping), utilitzar altres algorismes com k-NN (k-nearest neighbour) en comptes de GMMs, etc. però durant 40 anys cap nova aproximació completament diferent ha sabut resoldre aquest problema més efectivament. Així doncs, la millora en els sistemes de reconeixement de veu ha estat degut al refinament dels algorismes utilitzats i sobretot a l'augment de la capacitat de càlcul permetent entrenar als sistemes de ASR amb milers d'hores i podent obtenir resultats més ajustats.

Investigacions recents dutes a terme a IBM, Microsoft, Google i la universitat de Toronto en aquest camp utilitzant xarxes neuronals profundes (que fins ara no havien estat capaces de superar als mètodes basats en GMM-HMM) han estat molt esperançadores. Aquestes investigacions substitueixen els GMMs per xarxes neuronals profundes (DNNs: Deep neuronal networks) aconseguint millores significatives.

## **3.2. Processament del llenguatge natural**

El processament del llenguatge natural consisteix en la interacció entre persona i màquines mitjançant el llenguatge natural, és a dir, el llenguatge normal que utilitzen les persones per comunicar-se entre elles.

El processament del llenguatge natural és un problema difícil de resoldre computacionalment i per aquest motiu és un àrea d'estudi de la intel·ligència artificial.

### **3.2.1. Mètodes al llarg de la historia.**

Els primers intents de resoldre el problema utilitzaven ontologies que modelaven les estructures sintàctiques, quins tipus de paraules existeixen i com estan formades les paraules. Aquests programes desenvolupats en la dècada del 70s es basaven en un gran conjunt de regles escrites a mà. Aquests sistemes tenien problemes a l'hora de tractar amb entrades per a les quals no estaven preparats o eren poc usuals. Tot i així, funcionaven parcialment i van



servir per desenvolupar nombrosos programes amb els que una persona podia comunicar-se i rebre respostes amb un cert sentit (l'equivalent actual serien programes com Cleverbot).

A partir dels anys 80 però es va passar a fer servir un altre sistema. L'aprenentatge automàtic. Els sistemes basats en aquest sistema utilitzen algorismes generals d'aprenentatge, basats sovint en la inferència estadística per aprendre les regles del llenguatge automàticament a mesura que van rebent com a entrada nous textos. La implementació d'aquests sistemes ha anat evolucionant al llarg del temps (a mesura que s'han anat millorant les tècniques de Intel·ligència Artificial) des de Arbres decisionals (encara molt semblants als sistemes de regles escrites a mà anteriors) fins a sistemes totalment probabilístics (molt més avançats i amb un funcionament molt millor). Aquests sistemes donen resultats superiors pels següents motius:

El sistema automàticament aprèn les regles més generals sense necessitat de que un humà necessiti trobar-les prèviament i indicar-li (això facilita el treball i augmenta la resistència a errors).

La tècnica d'aprenentatge per inferència estadística fa que els models que es crea el programa siguin resistents a entrades poc comunes o errònies.

El sistema millora automàticament a mesura que rep més entrades, en canvi un sistema escrit a mà només pot millorar si un humà es dedica a modificar i augmentar la complexitat de les regles (el que provoca errors amb facilitat a més de ser un procés llarg i costós).

### **3.2.2. Question Answering**

L'aspecte més important del processament del llenguatge natural en els assistents personals és la resposta a preguntes (Question Answering). Un camp que en combina d'altres de la intel·ligència artificial com la cerca d'informació i la mineria de dades.

Els sistemes de Question Answering moderns utilitzen documents de text com a font de coneixement. També s'utilitza la web com a font de coneixement general. Els sistemes QA actuals típicament inclouen un mòdul que classifica les preguntes i determina el tipus de pregunta. Després de que la pregunta hagi sigut analitzada, el sistema aplica diversos mòduls a basats en NLP cada cop més complexos a parts reduïdes del text. És a dir, redueix el text en parts que són processats per separat. Llavors un altre mòdul utilitza motors de cerca per trobar paràgrafs que probablement contenen la resposta. Llavors un filtre selecciona fragments de text més petites i finalment un darrer mòdul extrau la resposta dels fragments seleccionats que creu que és més correcta.

## **3.3. Diferències i detalls i implementació de les tecnologies de cada un**

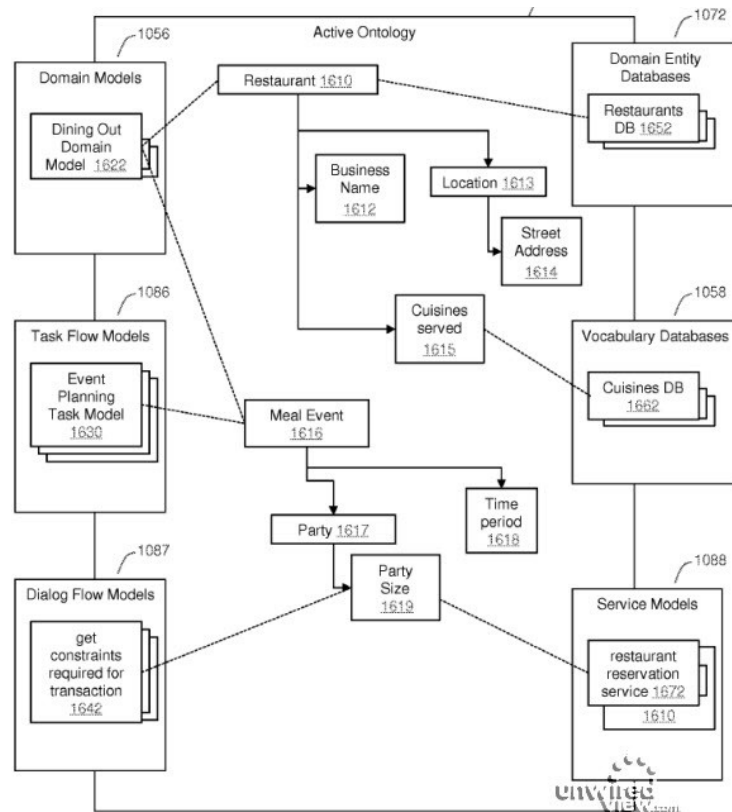
### **3.3.1 Siri:**

Siri consta de tres parts un reconeixedor de veu, un analitzador de gramàtica o processador de llenguatge natural i finalment un conjunt de proveïdors de serveis.

#### Ontologia simplificada

La diferència principal entre Siri i altres intents en la IA, és que Siri ha sigut pragmàtic: En comptes d'intentar abarcar tot el coneixement com s'havia intentat fer fins ara el que fa és basar-se només en les parts que creu que els usuaris poden estar interessats, com poden ser

restaurants, esports, esdeveniments, el temps, pel·lícules, viatges, hotels... Ho anomenen àrees especialitzades. Bàsicament el que fa és entendre certes paraules clau que creu que són importants i sobre les quals té informació i ho combina amb altres serveis web, bases de dades o regles d'interacció amb la màquina, també pots preguntar per informació general, en aquest cas, utilitza Wolfram Alpha per respondre o Wikipedia. En cas contrari, si preguntes a Siri sobre un domini que no coneix et respondrà amablement que no ho sap.



### Siri és "Sociable"

Un dels punts clau de Siri segons alguns experts és que Siri no vol fer quedar l'usuari com un ignorant i no intenta fer les recomanacions sense ser preguntat, també respon que no sap contestar-te quan es demana alguna cosa que no és capaç de respondre. A més, Siri porta un domini que és capaç de captar les bromes i respondre amb bromes també.#

### La ontologia es va ampliant (o està previst)

Apple vol que la seva tecnologia vagi més enllà dels mòbils. S'està pensant en exportar aquest sistema a cotxes, Macs, TVs i molts altres aparells de la vida quotidiana. Com que els temes que Siri ha de poder cobrir en cada un d'aquests casos són diferents s'haurà d'ampliar la ontologia per estendre-la a noves àrees o completar les que ja cobreix. Alguns exemples poden ser: Serveis locals (cines restaurants hotels...), interacció social (esdeveniments, compartir fotos, musica, missatges), comerç electrònic, navegació, cerques a Internet, informació meteorològica, tràfic, comunicació (missatgeria instantània, e-mail), interacció amb la casa (domòtica en general)...

Malgrat que no ha estat confirmat, es tenen indicis clars que l'empresa que controla els servidors a on arriben les peticions de Siri és Nuance. Nuance és una empresa dedicada al reconeixement de veu i d'imatge.

Com ha utilitzat les tècniques i quin benefici n'ha obtingut.

Utilitzar aquest plantejament d'ontologies especialitzades ha permès que Siri sigui capaç de contestar correctament a l'usuari i donar-li informació rellevant en la gran majoria dels casos, cosa que sistemes anteriors no eren capaços d'aconseguir (ja que les ontologies eren massa complexes i per tant era poc eficaç buscar informació, o bé eren més senzilles però no tenien suficient informació en conjunt).

El reconeixement de veu funciona mitjançant la tècnica de GMM-HMM, la qual és molt coneguda a l'àmbit de la IA ja que fins ara ha estat la que ha donat millors resultats (cosa que està canviant amb els nous models DNN-HMM).

### **3.3.2 Google Now**

Google ha creat unes eines de reconeixement de veu innovadores per fer-les servir amb Google Now. Ha començat a utilitzar sistemes basats en DNN-HMM, el qual li permet fer un reconeixement molt més precís.

Per tal de donar resposta a les preguntes i cerques de l'usuari Google ha creat l'anomenat Google Knowledge Graph. Aquest és una espècie d'ontologia o frame que conté 500 milions d'entrades. La part més innovadora d'aquest frame és Freebase, una base de coneixement que pot ser editada pels usuaris (semblant a altres serveis com Wikipedia). Freebase organitza la informació de forma semàntica (això vol dir organitzar la informació de manera que les màquines puguin entendre el significat d'aquesta de manera semblant a com ho faria un humà, i per tant poden realitzar operacions molt més complexes).

Encara que Freebase és un sistema innovador i amb molts seguidors que està tenint bastant èxit com a element del Google Knowledge Graph hi ha alguns aspectes que són criticats:

Freebase importa dades en massa, però les eines per transformar-les al format correcte i crear les relacions necessàries dins de la base del coneixement són massa complicades per ser utilitzades pel públic, per tant aquestes importacions només es fan de manera interna.

Encara que Freebase suporta traduccions el nom dels tipus sempre és en anglès, això comporta problemes a l'hora de crear un esquema universal.

Freebase no té cap solució per a valors desconeguts o no disponibles, per tant per a una cerca com "Mort per causes desconegudes", la cerca buscaria morts amb causes desconegudes, tant si les causes realment no es coneixen com si aquests valors són coneguts però no existeixen a la ontologia.

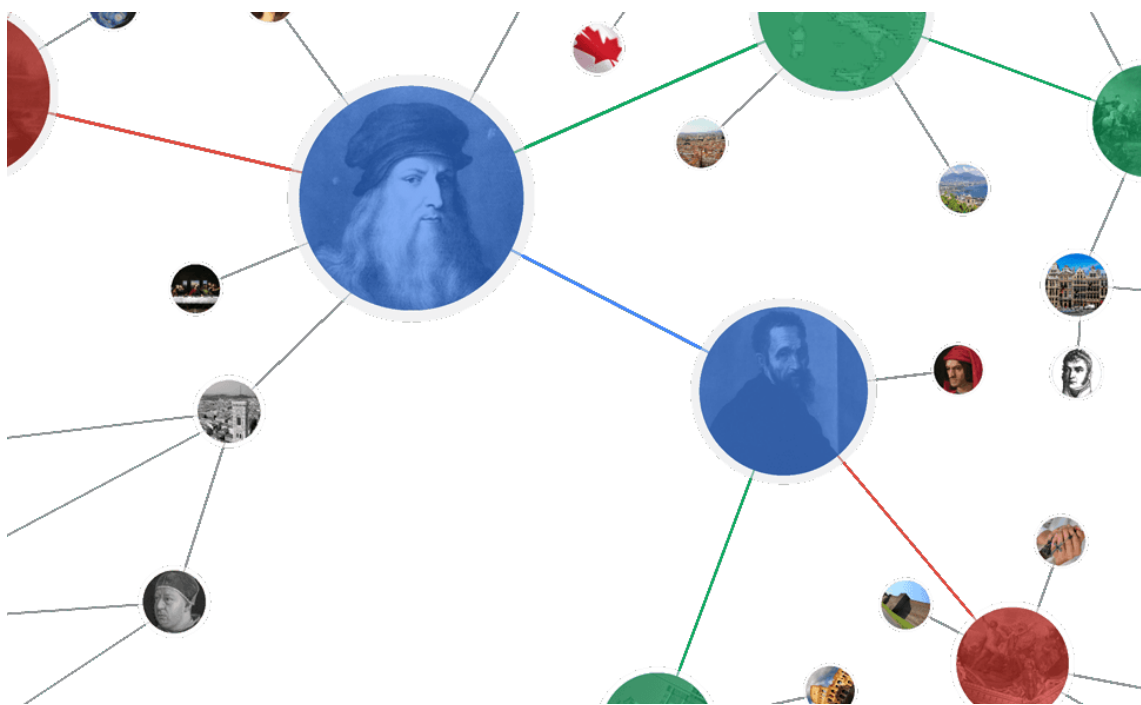
A més Google Now és capaç de recollir informació del usuari gràcies a que és capaç d'interaccionar amb els altres serveis de Google: pot llegir la bústia d'email, obtenir en tot moment la localització, i consultar les cerques que l'usuari ha fet a Google per tal de recollir dades i adaptar-se a les preferències i necessitats de l'usuari.

Com ha utilitzat les tècniques i quin benefici n'ha obtingut.

L'ús de sistemes innovadors en el reconeixement de veu permet a Google Now entendre a l'usuari amb una facilitat i precisió molt més alta, permetent que l'esforç que l'usuari ha de fer perquè el sistema l'entengui es redueixi, això fa que el servei sigui molt més accessible al públic. A més l'ús d'ontologies com el seu Google Knowledge Graph permet al sistema arribar a resultats per a les cerques de l'usuari i a més aconseguir que aquests siguin rellevants. Per últim la capacitat que té el sistema per aprendre, una tècnica estudiada també per la intel·ligència artificial, permet que Google Now pugui recopilar molta informació sobre l'usuari i aconseguir amb ella adaptar-se a l'usuari i arribar a poder donar-li informació fins i tot abans de que la demani, l'important d'aquest punt, és que gràcies a les tècniques d'intel·ligència artificial utilitzades aquesta informació és rellevant i realment útil a l'usuari (no només apareix perquè en general té uns certs interessos que va indicar al començar a utilitzar l'aplicació, per exemple), el que desmarca aquest producte de intents anteriors de realitzar aquestes tasques.

### 3.4 Limitacions de la tecnologia

Cal tenir en compte però, que per la seva naturalesa, aquesta tecnologia és imprecisa. Això implica que és possible que els resultats d'una execució no siguin els esperats. Per aquest motiu els fabricants han hagut de preveure mecanismes de recuperació dels errors de manera eficient. Per exemple: Si Google Now prediu que un cert contingut serà interessant per al usuari, però resulta que no és així, llavors l'usuari té la possibilitat de descartar aquest resultat. Llavors passen dues coses: per una banda es mostren al usuari una sèrie de resultats alternatius d'una forma que li pugui resultar familiar, això és, que els resultats es mostren de la mateixa manera que es mostren els resultats en el buscador de Google. D'aquesta manera es produeix una recuperació ràpida del error i l'usuari no en resulta tant perjudicat. Per altre banda Knowledge Graph actualitza les preferències del usuari i aprèn, de manera que la següent vegada pugui mostrar uns resultats diferents que puguin aconseguir l'usuari.



## **4 Impacte**

En aquest apartat s'analitza l'impacte que han tingut aquests productes des del punt de vista social, econòmic i empresarial.

### **4.1. Perquè Google i Apple estan interessats en tenir els seus propis i no deixen que siguin aplicacions externes**

Hi ha diferents motius que ens poden fer pensar perquè Google i Apple volen controlar aquests tipus d'aplicacions.

De fet, explícitament ells no prohibeixen que hi pugui haver aplicacions que facin aquesta funció (Siri va començar sent una app independent), però el que passa és que si mantenen una aplicació així, ja desanimen els possibles desenvolupadors que no s'atreviran a fer la competència directa a Google o Apple ja que tenen molt més nom i els clients seria difícil que els escollissin.

A més, en el cas d'Apple no es pot utilitzar la mateixa font de dades, a diferència de Google que permet l'ús de Google Knowledge Graph.

Un altre motiu és econòmic i de doble resant: D'una banda els assistents personals són un nou segment de mercat encara per explotar que pot donar grans beneficis a qui el controla, si una nova empresa arribés a dominar aquest segment, podria esdevenir una empresa molt potent i un competidor directe. Per altra banda, una aplicació així és un complement perfecte per un sistema operatiu mòbil com el que ells ja tenen i servirà per donar més vendes dels respectius dispositius Android i iOS.

Un altre motiu és la necessitat de tenir grans infraestructures per a poder emmagatzemar la informació i també per a processar-la. Això fa que poques empreses tinguin realment a l'abast poder fer una aplicació d'aquestes característiques amb un potencial real.

### **4.2. Acceptació**

Aquests productes encara que relativament nous (encara que versions més primitives ja existien abans) el públic ha mostrat gran interès per ells. Això és probablement degut a que per primer cop aquesta tecnologia de reconeixement de veu i comunicació amb el dispositiu funciona suficientment bé com per a mostrar bons resultats, i és un tipus de tecnologia que impressiona. A més obren tot un nou món de possibilitats en la interacció de les persones amb els seus dispositius que la gent vol explorar. Un altre motiu de la gran acceptació d'aquests productes és el fet de que els seus desenvolupadors són empreses molt conegudes amb una gran base de seguidors i la plataforma en la qual s'està desenvolupant és una de les més utilitzades a nivell mundial.

### **4.3. Que aporten aquests productes**

En aquest apartat s'analitzen els diferents beneficis i també riscos que aporten aquests productes tant per a les empreses com per els consumidors.

#### **4.3.1 per la empresa (riscos i beneficis)**

##### Beneficis

Un dels beneficis més evidents i immediats que té l'aparició d'aquestes noves tecnologies és l'impacte mediàtic que tenen, sobretot en els sectors que no pertanyen al món de la intel·ligència artificial (i de la informàtica en general) i que veuen aquests productes com una innovació extraordinària. Aquests productes són el fruit d'anys d'investigació i l'evolució de moltes tècniques que s'han anat perfeccionant, encara que el gran públic no sol pensar en aquest fet i pensa que aquests productes han estat desenvolupats en poc temps per empreses com Google o Apple, el que fa que l'estima que té la gent per aquestes empreses augmenti encara més. Però a aquestes empreses ja els va bé una publicitat així per tal de donar-se a conèixer i fer-se un nom dins del món de la informàtica de consum. En aquest aspecte Apple domina més. L'iPhone s'ha convertit en un producte gairebé de culte, això ha portat a que tingui un gran grup de seguidors i un grup igualment gran de detractors, aquesta petita guerra ha fet que el producte passi a ser molt conegut també entre altres grups del públic menys interessats en el tema. Indirectament aquest fet també ha donat bastant popularitat a Siri. Malgrat que Google no és tan bona publicitant els seus productes, molts dels detractors d'Apple i els seus productes han trobat refugi a Google i el seu sistema operatiu Android per a dispositius mòbils, el que ha fet que Google no es quedi enrere en aquest sector.

Tot i que pot semblar que aquestes empreses només desenvolupen aquests productes per publicitar-se, és clar que a llarg termini, el desenvolupament de noves tècniques en intel·ligència artificial és útil per a la comunitat en general.

Finalment cal destacar que aquests productes serviran de base per a la creació de noves innovacions en aquest camp i també suposaran la obertura de nous mercats dins de la informàtica i fins i tot fora del mateix camp de la informàtica.

##### Riscos

Per altre banda existeixen una sèrie de riscos que no tenen a veure en la tecnologia en si, sinó que són més de caire legal, social o fins i tot ètic.

A part de les malauradament habituals denúncies per violació de patents que hi ha en la particular guerra de patents que tenen els gegants de la informàtica. En aquests casos també són habituals les denúncies aquest cop per part de governs i organitzacions de consumidors. Ja que aquests productes deixen palesa la monitorització que fan les empreses als seus clients i espanten a molts dels defensors de la privacitat. Aquests últims tenen motius per estar descontents ja que malgrat les lleis de protecció de dades vigents, aquestes empreses acostumen a ser de països amb altres legislacions on el seu propi govern pot demanar informació sobre una persona i les empreses tenen la obligació de donar-li.

Un altre risc que corren aquestes grans empreses al presentar un producte punter és que aquest producte no tingui l'èxit esperat o que sigui un fracàs degut a que l'aplicació no està ven depurada. En aquests casos el nom de l'empresa sofreix un cop important degut a que el producte havia generat unes expectatives per sobre del resultat obtingut. Llavors aquest producte passa a ser la riota de la comunitat de consumidors i s'ha d'acabar retirant perquè ja és irrecuperable.

#### **4.3.2 Per a la gent que l'utilitza (riscos i beneficis)**

##### Beneficis

L'objectiu d'aquestes aplicacions és fer l'accés a la informació més ràpid i eficaç, de manera que intenten organitzar-la de manera que pugui resultar útil a l'usuari, l'altre cavall de batalla és intentar fer que la interacció no sigui utilitzant un teclat físic sinó amb la parla. Clarament si aquests productes aconseguixen l'objectiu que prometen suposarà un gran avenç en la pròpia interacció persona-màquina i podria suposar una revolució tal que deixéssim la època actual on tenim accés a una gran quantitat de informació, però que és una informació que no podem assimilar, per una època on la informació arribi de manera ordenada i quan la necessitem. És clar, que aquests productes poden ser una revolució.

##### Riscos

L'altre cara de la moneda són els riscos que l'ús d'aquesta tecnologia pot comportar.

El primer és el fet que aquests aparells poden donar una informació errònia que desorienti l'usuari, per tant, no s'haurien d'utilitzar en situacions crítiques ja que encara no són productes prou precisos com per ser utilitzats en aquesta finalitat, per exemple, Siri no sempre entén el que li dius i a vegades ho interpreta de forma incorrecta.

La part més crítica i el motiu per al qual aquests productes són més criticats és el fet que suposen una invasió flagrant a la intimitat de l'usuari. Això és degut a que per funcionar correctament necessiten grans quantitats de dades. Google Now per exemple s'intenta anticipar a l'usuari mostrant-li informació que pugui necessitar, per fer això, prèviament ha hagut de processar els correus electrònics de la persona, la seva localització geogràfica, té en compte un historial de cerques etc. Òbviament aquestes dades no són guardades al mateix dispositiu client des d'on s'executa l'aplicació sinó que es guarda en uns servidors que molt probablement estan en un altre país. Aquest fet, fa molt difícil que un usuari qualsevol pugui demanar a l'empresa que esborri les seves dades dels seus servidors (dret a l'oblit) quan vol deixar de ser-ne client. A més l'empresa segurament utilitzarà les dades per a altres finalitats que no són les desitjades per l'usuari quan aquest va entrar al servei i va acceptar uns termes i condicions sense llegir, com pot ser per finalitats publicitàries o vendre la informació a tercers, inclús els governs poden accedir a aquesta informació per investigar en contra de la voluntat de la pròpia empresa. Per tant, l'usuari ha d'anar alerta amb la informació que dona sobre ell a Internet, però aquests productes descrits són l'antítesi d'això ja que la informació flueix per totes bandes, encara que l'usuari no vulgui.

Aquest darrer problema fa pensar fins a on podem deixar arribar les empreses en aquest sentit, ja que es plantegen diversos problemes de caire ètic i moral que costaran molts anys de resoldre si és que mai s'arriben a resoldre.

## **5.Conclusions**

Es pot dir que els anys 2011-2012 han sigut els anys de l'assistent personal. Encara que és una tecnologia que s'està aplicant de forma comercial amb bons resultats des de fa poc, el públic sembla interessat i entusiasta amb ella. Això fa pensar que les empreses i grups de recerca es

dedicaran a millorar notablement aquesta tecnologia en els anys vinents i a crear les seves pròpies versions (per exemple Samsung ha desenvolupat l'”S Voice”, Nuance ha creat “Nina”, i Microsoft també està treballant per crear el seu propi sistema d'assistent personal). Tenint aquest fet en compte i afegint-hi el fet que la intel·ligència artificial (base d'aquests sistemes) en general està vivint grans avenços en molts dels seus àmbits podem esperar que els assistents personals passin a ser un producte de qualitat utilitzat en molts àmbits i no només els dispositius mòbils (podrien ser útils en hospitals, en PCs de sobretaula, i fins i tot a les futures cases domòtiques).

Però com tota nova tecnologia, els assistents personals porten alguns problemes. En aquest cas afecta a la privacitat. Els assistents personals recullen dades de tot el que l'usuari realitza i envia molta informació a servidors externs mitjançant Internet. En el futur s'haurà d'assegurar un bon control sobre la informació recollida per aquests assistents i com la manipulen.

## **6. Bibliografia**

Transparències del curs EECS E6870: Speech Recognition, impartit per Bhuvana Ramabhadran (IBM), Michael Picheny (IBM) i Stanley F. Chen (Columbia University) a la Columbia University:

<http://www.ee.columbia.edu/~stanchen/fall12/e6870/outline.html>

Transparències del curs CSCE 689-604: Special Topics in Speech Processing, impartit a la Texas A&M University impartit per Ricardo Gutierrez-Osuna (Primavera 2011):

[http://courses.cs.tamu.edu/rgutier/csce689\\_s11/](http://courses.cs.tamu.edu/rgutier/csce689_s11/)

Curs de Coursera impartit per Dan Jurafsky i Christopher Manning de la Stanford University:

<https://www.coursera.org/course/nlp>

Geoffrey Hinton, Andrew Senior, Vincent Vanhoucke, Abdel-rahman Mohamed, Navdeep, Patrick Nguyen, Tara Sainath, and Brian Kingsbury “Paper Deep Neural Networks for Acoustic Modeling in Speech Recognition” Novembre 2012 IEEE SIGNAL PROCESSING MAGAZINE

<http://www.cs.toronto.edu/~hinton/absps/DNN-2012-proof.pdf>

Edicions digitals de la IEEE Signal Processing Magazine:

<http://www.signalprocessingsociety.org/publications/electronic-pubs/>

Vídeo Sobre Gauss Mixture Model

<http://www.youtube.com/watch?v=Rkl30Fr2S38>

Curs de Coursera sobre Xarxes Neuronals impartit per Geoffrey Hinton Universitat de Toronto:

[https://class.coursera.org/neuralnets-2012-001/lecture/download.mp4?lecture\\_id=4](https://class.coursera.org/neuralnets-2012-001/lecture/download.mp4?lecture_id=4)

Article sobre com funciona Siri: Staska 12 d'octubre de 2011



<http://www.unwiredview.com/2011/10/12/how-siri-on-iphone-4s-works-and-why-it%E2%80%99s-a-big-deal-apple%E2%80%99s-ai-tech-details-in-230-pages-of-patent-app/>

Ronan Collobert Jason Weston "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning"

[http://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&ved=0CEIQFjAB&url=http%3A%2F%2Ftrac.assembla.com%2Fchatbot\\_tesis%2Fexport%2F120%2Ftrunk%2Fmaterial%2Fpapers%2Funified\\_nlp.pdf&ei=WajMUP3wGoezhAelioHwBQ&usg=AFQjCNE0qyLY-nqgq1qhjJN2LyAAJtkx9g&sig2=46VjmwAgRaO\\_2jE9DA8pdQ](http://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&ved=0CEIQFjAB&url=http%3A%2F%2Ftrac.assembla.com%2Fchatbot_tesis%2Fexport%2F120%2Ftrunk%2Fmaterial%2Fpapers%2Funified_nlp.pdf&ei=WajMUP3wGoezhAelioHwBQ&usg=AFQjCNE0qyLY-nqgq1qhjJN2LyAAJtkx9g&sig2=46VjmwAgRaO_2jE9DA8pdQ)

Article de Janie Chang sobre els avenços en reconeixement de veu. 29 d'agost de 2011

<http://research.microsoft.com/en-us/news/features/speechrecognition-082911.aspx>

Vídeo de presentació de Siri (quan encara no ho havia comprat Apple). 5 de febrer de 2010

<http://tomgruber.org/siri-launch>

Article de Will Knight sobre la sociabilitat de Siri. 24 d'abril de 2012

<http://www.technologyreview.com/review/427664/social-intelligence/>

Article de Tom Simonite sobre Google Now . 28 de setembre de 2012

<http://www.technologyreview.com/news/429345/googles-answer-to-siri-thinks-ahead/>

Article de James Johnson sobre els llenguatges que suporta Google Now. Agost de 2012:

<http://www.inquisitr.com/304873/google-voice-search-now-supports-13-more-languages/>

Article de Tom Simonite sobre el Knowledge Graph. 14 de Juny de 2012

<http://www.technologyreview.com/news/428180/googles-new-brain-could-have-a-big-impact/>

Article sobre com funciona Siri. 5 d'octubre de 2012

<http://www.jeffwofford.com/?p=817>

Entrada a la Wikipedia del processament del llenguatge natural

[http://en.wikipedia.org/wiki/Natural\\_language\\_processing](http://en.wikipedia.org/wiki/Natural_language_processing)

Entrada a la Wikipedia de Google Now:

[http://en.wikipedia.org/wiki/Google\\_Now](http://en.wikipedia.org/wiki/Google_Now)

Entrada a la Wikipedia sobre Question Answering:

[http://en.wikipedia.org/wiki/Question\\_answering](http://en.wikipedia.org/wiki/Question_answering)

Entrada a la Wikipedia de Siri:

[http://en.wikipedia.org/wiki/Siri\\_%28software%29](http://en.wikipedia.org/wiki/Siri_%28software%29)

Entrada a la Wikipedia de Freebase

<http://en.wikipedia.org/wiki/Freebase>

Entrada a la Wikipedia sobre els assistents personals intel·ligents:

[http://en.wikipedia.org/wiki/Intelligent\\_personal\\_assistant](http://en.wikipedia.org/wiki/Intelligent_personal_assistant)

Entrada a la Wikipedia sobre freqüències de mostreig:

[http://en.wikipedia.org/wiki/Sampling\\_frequency](http://en.wikipedia.org/wiki/Sampling_frequency)

Historia sobre el reconeixement de veu:

[http://www.pcworld.com/article/243060/speech\\_recognition\\_through\\_the\\_decades\\_how\\_we\\_ended\\_up\\_with\\_siri.html](http://www.pcworld.com/article/243060/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html)