|  |  |
| --- | --- |
|  | **Arquitectura de Aplicaciones**  **Cuatrimestre 1 de 2020** |

**TRABAJO PRÁCTICO DE INVESTIGACIÓN**

**Watson**

**Aula Nº: 106 - Turno: Miércoles, noche**

**GRUPO Nº: 2**

**1086664 Ginabreda, Santiago**

**1087295 Gutierrez Bertelli, Carolina**

**1021972 Nocetti, Alan**

**1052130 Corvetto, Dana**

**1045775 Covre, Juan Martin**

**1056461 Marino, Damián**

**Fecha de presentación: 01/06/2020** **Fecha de devolución \_\_\_/\_\_\_/\_\_\_\_\_\_**

**Observaciones**

………………………………………………………………………………………………..

………………………………………………………………………………………………..

…………………………………………………………………………………………………

…………………………………………………………………………………………………

…………………………………………………………………………………………………

# 

[**Qué es Watson?**](#_v16b43c8i7x0) **2**

[DeepQA IBM](#_u7glvfg4ykkg) 3

[Power7](#_45kkddxbchju) 4

[Principal Uso](#_1b39spsc8zak) 5

[Funcionamiento](#_c3hu826nibjz) 7

[1- Adquisición de contenido](#_zdwqpbmi1fhq) 7

[2- Análisis de preguntas](#_w5sllruzyhqf) 8

[3- Gestión de Hipotesis](#_n380rfobl6y6) 8

[4- Filtrado Suave](#_s3nzgy985ii0) 9

[5- Puntuación de Hipótesis y Evidencias](#_tl3ofom89hpd) 9

[6- Mezcla final y clasificación](#_pzz19w3quis6) 10

[Ejemplo](#_1us5on3ol11r) 10

[Speech to text](#_gcujj4txz2dg) 10

[Visual recognition](#_7xp24sb5us5j) 11

# 

# Qué es Watson?

**Watson** es un sistema optimizado de cargas de trabajo basado en arquitectura IBM DeepQA la cual se ejecuta en un clúster de servidores basados en el procesador IBM® POWER7®. Después de cuatro años de investigación y desarrollo intensos por un equipo de investigadores de IBM, Watson compitió en Jeopardy! en febrero de 2011, rindiendo al nivel de expertos humanos en términos de precisión, confianza y velocidad contra dos de los Campeones de Jeopardy! más conocidos y exitosos, Ken Jennings y Brad Rutter

En 1997, Deep Blue, el sistema de juego de ajedrez computarizado desarrollado por IBM Research, captó la atención mundial al competir con éxito contra el campeón en ajedrez Gary Kasparov. Fue la culminación de un gran desafío para lograr avances en la ciencia de la computación en una forma que captó mucho interés popular.

Hoy en día, las compañías están capturando cada vez más información fundamental de negocios en documentación en lenguaje natural, por lo que hay un creciente interés en sistemas optimizados de cargas de trabajo que analizan a fondo el contenido de preguntas en lenguaje natural para responder esas preguntas con precisión. Los avances en la tecnología de respuesta a preguntas (QA) ayudarán cada vez más a los profesionales a tomar decisiones fundamentales y oportunas en áreas como la asistencia médica, la inteligencia empresarial, el descubrimiento de conocimientos, la gestión de conocimientos empresariales y la asistencia al cliente.

Con QA en mente, IBM se propuso un desafío para construir un sistema computacional denominado “Watson” (en honor a Thomas J. Watson, fundador de IBM), el cual pudo competir al nivel de campeones humanos en tiempo real en el programa de preguntas de la TV Americana Jeopardy!

El software incorporado en Watson fue escrito utilizando los lenguajes de programación Java y C++, y usa dos estructuras empleadas por la Apache Software Foundation — la estructura Hadoop para computación distribuida, y el sistema UIMA (Unstructured Information Management Architecture) para análisis de materiales impresos — así como el software DeepQA desarrollado por IBM y el sistema operativo SUSE Linux Enterprise Server 11.

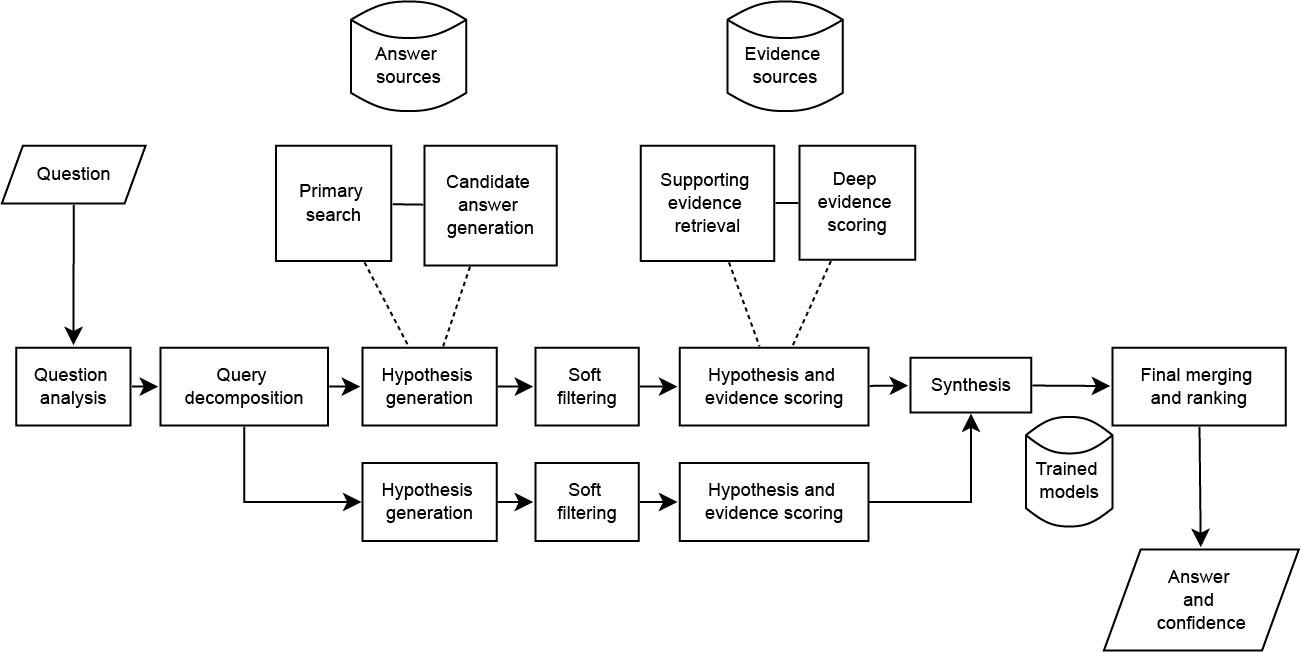
Watson representa un impresionante avance en el diseño y la analítica de sistemas. Ejecuta la tecnología **DeepQA** de IBM, un nuevo tipo de capacidad analítica que puede realizar miles de tareas simultáneas en segundos para brindar respuestas precisas a las preguntas.

Watson funciona con tecnología del procesador **IBM POWER7**, y es un ejemplo de las complejas cargas de trabajo analíticas que son cada vez más comunes y fundamentales para el éxito y la competitividad empresariales en el entorno de muchos datos de hoy en día.

## 

## 

## DeepQA IBM

**

*Arquitectura de alto nivel empleada por DeepQA*

DeepQA es una arquitectura basada en evidencia probabilística masivamente paralela. Para el Desafío de Jeopardy!, se utilizan más de 100 técnicas diferentes para analizar el lenguaje natural, identificar fuentes, encontrar y generar hipótesis, buscar y marcar evidencia, y unir y clasificar hipótesis. Mucho más importante que cualquier otra técnica en particular es la forma en que se combinan todas estas técnicas en DeepQA, de forma que los enfoques superpuestos puedan aportar sus cualidades y contribuir a mejoras en precisión, confianza o velocidad.

DeepQA es una arquitectura que viene junto con una metodología, aunque no es específica para el Desafío Jeopardy! IBM comenzó a adaptarla a diferentes aplicaciones empresariales y problemas de desafíos de exploración, incluida la medicina, la investigación empresarial y el juego.

Los principios globales en DeepQA son:

* Paralelismo masivo: aprovecha el paralelismo masivo en la consideración de interpretaciones e hipótesis múltiples.
* Muchos expertos: facilita la integración, aplicación y evaluación contextual de una amplia variedad de analíticas de preguntas y contenidos probabilísticos vagamente asociados.
* Valoración de confianza dominante: ningún componente único se compromete a una respuesta; todos los componentes producen funciones y confianzas relacionadas, marcando diferentes interpretaciones de preguntas y contenidos. Un sustrato subyacente de procesamiento de confidencias aprende a acumular y combinar los puntajes.
* Integración de conocimiento superficial y profundo: equilibra el uso de semántica estricta y semántica superficial, aprovechando ontologías vagamente formadas.

DeepQA se desarrolla utilizando Apache UIMA, una implementación de marco de la Arquitectura de Gestión de Información No Estructurada. UIMA fue diseñado para soportar la interoperabilidad y escalabilidad horizontal de aplicaciones de análisis multimodal y de texto. Todos los componentes en DeepQA se implementan como anotadores UIMA. Éstos son los componentes que analizan texto y producen anotaciones o afirmaciones acerca del texto. Con el tiempo, Watson evolucionó, por lo que ahora el sistema cuenta con cientos de componentes. UIMA facilitó una integración, prueba y evaluación rápidas de componentes.

Las primeras implementaciones de Watson se ejecutaron en un solo procesador, el cual necesitó de dos horas para responder a una sola pregunta. No obstante, la computación de DeepQA es muy paralela, por lo que puede dividirse en una cantidad de partes independientes, cada una de las cuales puede ejecutarse por un procesador por separado. UIMA-AS, el cual es parte de Apache UIMA, permite la escalabilidad horizontal de aplicaciones UIMA utilizando mensajería asíncrona. Watson utiliza UIMA-AS para obtener una escalabilidad horizontal en 2.800 núcleos POWER7 en un clúster de 90 servidores IBM Power® 750. UIMA\_AS gestiona toda su comunicación entre procesos, utilizando el estándar JMS abierto. La implementación de UIMA-AS en POWER7 le permitió a Watson proporcionar respuestas de uno a seis segundos.

Watson tiene aproximadamente 200 millones de páginas con contenido de lenguaje natural (equivalente a la lectura de 1 millón de libros). Watson utiliza el marco Apache Hadoop para facilitar el preprocesamiento de grandes volúmenes de datos para crear conjuntos de datos en la memoria utilizados durante el tiempo de ejecución. Los anotadores DeepQA UIMA de Watson fueron implementados como asignadores en la infraestructura map-reduce de Hadoop, la cual los distribuyó en todos los procesadores del clúster. Hadoop contribuye a un óptimo uso del CPU y también brinda herramientas convenientes para implementar, gestionar y monitorear el proceso de análisis de datos.

## Power7

Watson aprovecha el desempeño de procesamiento paralelo masivo de sus procesadores POWER7 para ejecutar sus miles de tareas de DeepQA simultáneamente en núcleos de procesadores individuales. Cada uno de los 90 servidores IBM Power 750 basados en clústeres cuenta con 32 núcleos POWER7 ejecutándose a 3.55 GHz. Los servidores ejecutan el sistema operativo Linux®, y se ubican en 10 bastidores junto con los nodos I/O y los hubs de comunicación relacionados.

El sistema cuenta con un total combinado de 16 Terabytes de memoria y puede funcionar a 80 Teraflops (trillones de operaciones por segundo).

Con su diseño innovador con procesador de 8 núcleos, POWER7 es ideal para un procesamiento masivamente paralelo de algoritmos analíticos de Watson. POWER7 también cuenta con 500 gigabytes con un ancho de banda de comunicaciones en chip, lo cual contribuye a una eficiencia excepcional de utilización tanto de la memoria como del procesador. Y debido a que cada servidor cuenta con 32 núcleos POWER7 de alto rendimiento con hasta 512 GB de memoria, el Power 750 es una plataforma ideal para el procesador de Watson y los procesos Java que necesitan de mucha memoria.

## 

## Principal Uso

La capacidad de Watson de comprender el significado y contexto del lenguaje humano, y de procesar información rápidamente para encontrar respuestas precisas a preguntas complejas, representa un enorme potencial para transformar la forma en que las computadoras pueden ayudar a las personas a cumplir tareas en los negocios y sus vidas personales.



*Cuadrante Mágico Gartner 2020 - Líderes en Plataformas Data Science y Machine Learning*

IBM es la única que se posiciona como “Challengers”. Los *«challengers»* se comportan los suficientemente bien en el mercado, como para en un momento determinado, convertirse en una amenaza seria para los líderes. Tienen productos sólidos y los recursos suficientes como para mantener un crecimiento continuo.

Sin embargo, de momento carecen del tamaño suficiente o de la influencia global que sí disfrutan de las empresas que se encuentran en el cuadrante *«leaders»*.

La intención de desarrollar las capacidades comerciales de Watson se plasmó en 2014, con la puesta en marcha del «**IBM Watson Group**» una nueva división de negocio encabezada por Michael Rodin y a la que comenzarán a prestar servicio más de 2.000 personas.

Desde su puesta en marcha, IBM ha invertido más de 1.000 millones de dólares en la división, que ha centrado sus esfuerzos en el desarrollo de tres grandes ramas de crecimiento: **Watson Discovery Advisor, Watson Engagement Advisor, y Watson Explorer**.

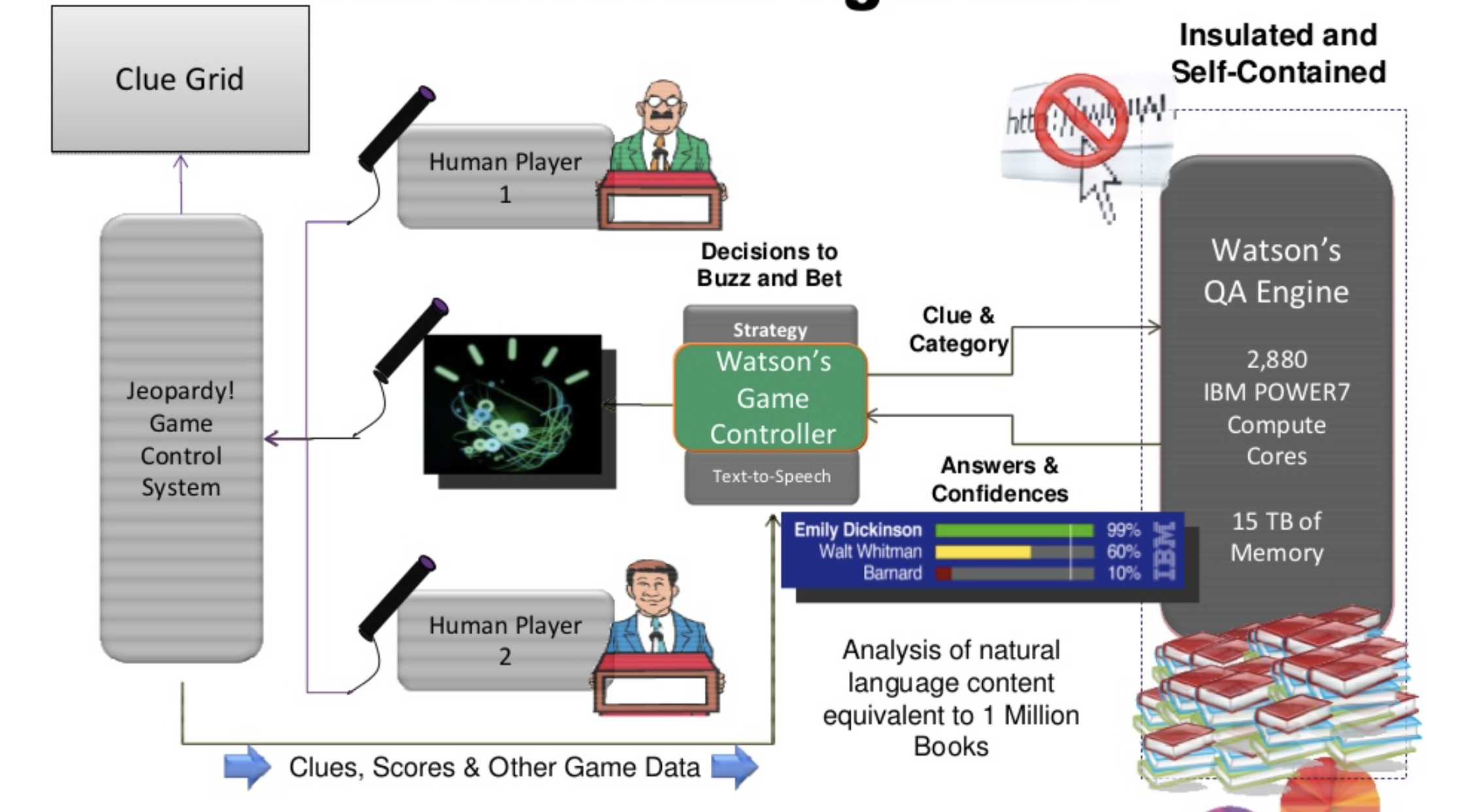
La primera, **Watson Discovery Advisor** trabaja en proyectos de investigación relación con la industria farmacéutica y la biotecnológica. **Watson Engagement Advisor** es el área desde el que se desarrollan aplicaciones self-service (chatbots, asistentes virtuales, etc.) para a partir de la interpretación del lenguaje natural, dar respuesta a las necesidades de los usuarios de empresa. Por último, **Watson Explorer** ayuda a los usuarios a descubrir, estructurar y compartir información a partir de enormes volúmenes de datos.

¿Y en qué proyectos ha estado involucrado últimamente? Sin ir más lejos, en el último US Open de tenis, Watson fue la plataforma sobre la que se construyó el asistente inteligente «Ask the US Open», que gracias a sus capacidades conversacionales informaba a los aficionados que acudían al torneo sobre todo lo que tenían que saber para mejorar su experiencia: consejos «antes de ir», puntos de interés en el estadio, integración de mapas, resultados y puntuaciones en directo, alertas meteorológicas, etc.

En el campo financiero, el banco italiano Creval ha conseguido que IBM Watson reduzca hasta en un 80% las peticiones de información que recibe su call center y que necesitan ser respondidas por un ser humano y al colaborar Thomson Reuters está consiguiendo que las empresas cumplan sin problemas con todas las regulaciones que se han puesto en marcha para gestionar la información y la privacidad de todo tipo de clientes, como en la aplicación efectiva de la GDPR).

Y no solo. En el campo experimental, ha desarrollado por ejemplo, una [«lengua electrónica» (IBM Hypertaste)](https://www.muycomputerpro.com/2019/08/06/hypertaste-lengua-electronica-ibm) que, inspirada en el funcionamiento del sentido del gusto humano, tiene como detectar y analizar distintos tipos de líquido rápidamente y sin necesidad de acudir a un laboratorio, o proyectos médicos tan interesantes como IBM Watson for Oncology, capaz de combinar la experiencia de los oncólogos en el cuidado del cáncer para ayudar a los médicos a considerar tratamientos individualizados para sus pacientes.

## Funcionamiento



### 1- Adquisición de contenido

Como una de las fases iniciales**,** la documentación del área de conocimiento en la que se va a entrenar el sistema cognitivo se “ingesta” en Watson. Dicha información contiene las respuestas y las evidencias en las que el sistema se va a basar a la hora de dar una respuesta. La selección de qué información es relevante y fiable la realiza un experto del área en concreto.

Las fuentes de Watson incluyen una amplia gama de enciclopedias, diccionarios, tesauros, artículos de servicio de noticias, obras literarias, y así sucesivamente. Dado un corpus de referencia razonable, a continuación DeepQA, aplica un proceso automático de expansión del corpus.

Posteriormente, Watson debe ser **entrenado por humanos** para aprender y entender contenido específico de diferentes sectores, especialidades e idiomas. En otras palabras, Watson aprende de una forma experta, no solo de manera general.

Esto implica capacitar al sistema para reconocer patrones complejos y abstractos, alimentándose con grandes cantidades de datos etiquetados y luego trabajando con expertos humanos para refinar las respuestas. A través de sucesivas rondas de información y retroalimentación, la comprensión y las respuestas de Watson mejoran. Watson aprende continuamente de las interacciones con personas y entornos.

### 2- Análisis de preguntas

El primer paso en tiempo de ejecución es el análisis de la pregunta, durante el cual se trata de comprender lo que se está preguntando y se realizan los análisis iniciales que determinan cómo será procesada la pregunta por el resto del sistema. El enfoque DeepQA en esta etapa se basa en una mezcla de expertos, y Watson produce análisis poco profundos, análisis profundos, formas lógicas, las etiquetas semánticas, correferencias, relaciones, entidades con nombre, y así sucesivamente.

**Clasificación de la pregunta**. Se identifican los tipos de preguntas o partes de las preguntas que requieren un procesamiento especial. Esto puede incluir cualquier cosa desde una sola palabra con significados potencialmente dobles hasta párrafos que pueden tener funcionalidades sintácticas, semánticas, retóricas.

**El enfoque y la detección de LAT.** Un tipo de respuesta léxica es una palabra o frase nominal en la pregunta que especifica el tipo de la respuesta sin ningún intento de entender su semántica.

**Detección de relación.** La mayoría de las preguntas contienen relaciones, ya sean predicados sintácticos sujeto-verbo-objeto o relaciones semánticas entre entidades. Watson utiliza la detección de relación a través de todo el proceso.

La capacidad actual de Watson para utilizar eficazmente bases de datos preparadas para simplemente buscar las respuestas se limita a menos del 2 por ciento de las pistas. Por ello hay que buscar en todo tipo de datos. Sólo en pocos casos Watson es capaz de encontrar las relaciones con facilidad, siendo esta un área de importante mejora.

**Descomposición**. DeepQA utiliza el análisis profundo, basado en normas y métodos estadísticos de clasificación tanto para reconocer si las preguntas deben ser descompuestas como para determinar la mejor forma de romperla en subpreguntas. Incluso si la pregunta no tenía por qué ser descompuesta para encontrar una respuesta, este método ayuda a mejorar la confianza global del sistema.

### 3- Gestión de Hipotesis

La generación de hipótesis toma los resultados de análisis las preguntas y produce respuestas candidatas mediante la búsqueda en las fuentes del sistema Cada respuesta candidata se coloca de nuevo en la pregunta y se considera una hipótesis, que el sistema tiene que demostrar correcta con cierto grado de confianza. Esta es la "de la búsqueda primaria" para distinguirla de las búsquedas realizadas posteriormente en la recopilación de pruebas.

**Búsqueda primaria.** El objetivo es encontrar tantas respuestas potenciales como sea posible sobre la base de los resultados del análisis de la pregunta. Hay que establecer un compromiso entre precisión y los recursos de computación utilizados. En el 85 por ciento de las preguntas, el sistema genera la respuesta correcta en algún lugar dentro de los 250 primeros candidatos.

**Generación de Respuesta Candidata.** Mediante distintas técnicas el sistema utiliza el resultado de la búsqueda para producir la respuesta candidata.

Este paso favorece significativamente el número más que la precisión, con la expectativa de que el resto del proceso encontrará la respuesta correcta, incluso si el conjunto de candidatos es muy grande. Uno de los objetivos del diseño del sistema, por lo tanto, es tolerar el ruido en las primeras etapas del proceso y elevar la precisión más adelante. Como consecuencia, Watson genera varios cientos de respuestas candidatas en esta etapa.

### 4- Filtrado Suave

Es la aplicación algoritmos de puntuación ligeros (con uso uso menos intensivo de los recursos) para podar desde un conjunto más amplio de candidatos iniciales hasta un conjunto más pequeño de los candidatos antes de aplicar los componentes de puntuación más intensivos Los candidatos que no pasan el umbral de filtrado se envían directamente a la etapa final.

### 5- Puntuación de Hipótesis y Evidencias

Las respuestas candidatas que pasan el umbral filtrado suave se someten a un riguroso proceso de evaluación que consiste en la recopilación de pruebas adicionales de apoyo para cada respuesta candidata o hipótesis,

**La recuperación de pruebas.** El sistema busca nuevas evidencias para la respuesta candidata.

**Puntuación**. El paso de puntuación es donde se lleva a cabo la mayor parte del análisis profundo de contenido. DeepQA fomenta la inclusión de muchos componentes diferentes, o puntuadores, que tienen en cuenta las diferentes dimensiones de las preguntas y producen una puntuación. DeepQA proporciona un formato común para que los puntuadores evalúen las hipótesis. Por ejemplo, Watson cuenta con más de 50 componentes de puntuación que producen distintos resultados.

Estos puntuadores tienen en cuenta cosas como el grado de correspondencia entre la estructura de un pasaje de predicado-argumento y la pregunta, la fiabilidad de la fuente del texto, la ubicación geoespacial, relaciones temporales, clasificación taxonómica, las relaciones léxicas y semánticas de las que se sabe que el candidato participa, en la correlación del candidato con los términos que la pregunta, su popularidad (o la oscuridad), sus alias, y así sucesivamente.

Cada uno de los puntuadores implementados, cómo funcionan, cómo interactúan, y su impacto independiente en el desempeño de Watson merecen su artículo de investigación propio.

Ningún algoritmo domina. La facilidad de DeepQA para incluir nuevos algoritmos y las herramientas representa una de las contribuciones más importantes y duraderas Watson.

### 6- Mezcla final y clasificación

Por último se mezclan todas las respuestas en una tabla final y se realiza un ranking de los cientos de respuestas candidatas. Watson ha adoptado un enfoque de aprendizaje automático que requiere usar el sistema a través de un conjunto de preguntas de entrenamiento con respuestas conocidas y la formación de un modelo basado en las puntuaciones.

## Ejemplo

Para el caso de estudio, decidimos realizar la prueba de dos productos dentro del catálogo de Watson.

Para realizarlo, se decidió utilizar Python, y se realizaron llamadas a los servicios provee la plataforma.

Se realizó el registro correspondiente en IBM Cloud, y la generación de las API Key para los servicios que se utilizaron.

### Speech to text

El servicio Speech to Text convierte la voz humana en palabras escritas. Se puede utilizar en cualquier lugar donde se necesite llenar el hueco entre la palabra hablada y la forma escrita, incluido el control de voz de los sistemas incorporados, la transcripción de reuniones y teleconferencias y el dictado de correo electrónico y notas. Este servicio fácil de usar utiliza la inteligencia artificial para combinar la información sobre la estructura de la gramática y los idiomas con los conocimientos de la composición de la señal de audio para generar una transcripción precisa.

En la prueba realizada se observa como el servicio transcribe el audio que se le envió en los parámetros correspondientes, donde originalmente el audio dice *“Hola, somos el grupo dos, y esto es una prueba para la materia arquitectura de aplicaciones”*.

[Audio test](https://drive.google.com/file/d/1XNvthUjWhU-Ztu5Rz1_5Dq-rtRr5CTr2/view?usp=sharing)

alnocetti@alnocetti-pc:~/vsc\_workspace/watson$ python3 speech-to.text.py

{

"results": [

{

"keywords\_result": {

"hola": [

{

"normalized\_text": "hola",

"start\_time": 0.88,

"confidence": 0.96,

"end\_time": 1.35

}

]

},

"alternatives": [

{

"confidence": 0.94,

"transcript": "hola somos del grupo dos y esto es una prueba para materia arquitectura de aplicaciones "

},

{

"transcript": "hola somos del grupo dos y esto es una prueba para la materia arquitectura de aplicaciones "

},

{

"transcript": "hola somos del grupo dos y esto es una prueba para materias arquitectura de aplicaciones "

}

],

"final": true

}

],

"result\_index": 0

}

### Visual recognition

IBM Watson™ Visual Recognition utiliza algoritmos de deep learning para identificar escenas, y objetos en imágenes que se envían al servicio. Te permite crear y entrenar tu propio cliente de clasificación, dependiendo de las necesidades de cada uno.

IBM Watson™ Visual Recognition cuenta con distintos modelos de clasificación

* **Modelo general**Genere palabras clave de clase que describan la imagen. Utilice sus propias imágenes o extraiga o los URL de imágenes desde páginas web de acceso público para su análisis.
* **Modelo de alimentos**Utilice un vocabulario especializado de más de 2000 alimentos para identificar comidas, productos alimentarios y platos con precisión mejorada.
* **Modelo personalizado**Cree clasificadores visuales personalizados y exclusivos. Utilice el servicio para reconocer los conceptos visuales personalizados no disponibles con el modelo general.
* **Modelo explícito**Evalúe si una imagen incluye contenido inapropiado o para adultos que puede no ser adecuado para el público en general.

En las pruebas realizadas se utilizó el “modelo general” y se le envió la imagen de un perro, y obtuvimos la siguiente respuesta.



alnocetti@alnocetti-pc:~/vsc\_workspace/watson$ python visual-recognition.py

visual-recognition.py:6: DeprecationWarning: watson-developer-cloud moved to ibm-watson. To get updates, use the new package. iam\_apikey='TFLKKzCDBUaC6JsRHmrDVrQWLbToHTwzRgkPsR5bGMSK')

{

"images": [

{

"classifiers": [

{

"classes": [

{

"score": 0.903,

"class": "beagling (dog)",

"type\_hierarchy": "/animal/domestic animal/dog/beagling (dog)"

},

{

"score": 0.919,

"class": "dog"

},

{

"score": 0.923,

"class": "domestic animal"

},

{

"score": 0.924,

"class": "animal"

},

{

"score": 0.955,

"class": "light brown color"

}

],

"classifier\_id": "default",

"name": "default"

}

],

"resolved\_url": "https://us.hola.com/imagenes/estar-bien/20190820147813/razas-perros-pequenos-parecen-grandes/0-711-550/razas-perro-pequenos-grandes-a.jpg",

"source\_url": "https://www.hola.com/imagenes/estar-bien/20190820147813/razas-perros-pequenos-parecen-grandes/0-711-550/razas-perro-pequenos-grandes-a.jpg"

}

],

"custom\_classes": 0,

"images\_processed": 1

}