

IBM Watson

Nora Schröter

24. Mai 2017

Inhaltsverzeichnis

0.1	Timeline	2
0.2	Wie funktioniert Watson	3
0.2.1	Was ist Jeopardy	3
0.2.2	DeepQA	5
0.2.3	Schwächen	8
0.3	Anwendungen	10
0.3.1	Generell	10
0.3.2	die Firma IBM Watson	10
0.3.3	Medizin	10
0.3.4	Olli derBus	11
0.3.5	Onlinepräsenz	11
0.3.6	Weitere	11

0.1 Timeline

- 1997 • DeepBlue gewinnt gegen Kasparov
- 2004 • Charles Lickel hat eine Idee für eine neue IBM Challenge
- 2005 • Paul Horn unterstützt das Projekt
- 2006 • Erste Testläufe ergeben eine 15% Erfolgsrate
- 2007 • Ferucci bekommt 3-5 Jahre und 15 Researcher zur Verfügung
- 2008 • Watson gewinnt vereinzelt gegen Champions
- 2010 • Es werden regelmäßig menschliche Kandidaten geschlagen
- 2011 • Watson gewinnt offiziell gegen ehemalige Champions

1997 gewinnt IBMs Schachcomputer DeepBlue gegen den Weltmeister Kasparov. Damit ist Schach als Entwicklerproblem "gelöst" und IBM begibt sich auf die Suche nach einer neuen Herausforderung.

2004 geht der IBM Entwickler Charles Lickel mit ein paar Kollegen essen. Irgendwann merken sie, dass das gesamte Restaurant in totenstille verfallen ist. Im Fernsehen läuft "Jeopardy!" wo gerade Ken Jennings seine 74-Spiele lange Gewinnreihe verteidigt. Lickel hat eine neue Idee für eine "IBM Challenge" gefunden

2005 findet er Unterstützung von IBM Research Executive Paul Horn, trotzdem haben die beiden Schwierigkeiten Entwickler zu finden, die die komplexe Aufgabe (immerhin ist es nicht so analytisch wie Schach) annehmen wollen. Kurz darauf nimmt der Entwickler David Ferucci das Angebot an.

Watsons Vorgänger, genannt Piquant, schaffte es ca. 35% der Hinweise richtig zu beantworten und brauchte für jeden mehrere Minuten. Im tatsächlichen Jeopardy Spiel hat Watson wenige Sekunden und muss 80% richtig beantworten. Die gefundenen Probleme galten als unlösbar.

2006 lässt Ferucci die ersten Testläufe mit Watson starten. Von den etwa 500 Hinweisen aus vergangenen Jeopardy-Folgen konnte Watson nur etwa 15% richtig beantworten. Seine menschlichen Kollegen dagegen "buzzten" zu etwa der Hälfte aller Fragen und lagen bei diesen 95% der Zeit richtig.

2007 Wird das Projekt stark vergrößert und IBM gibt dem Team 3-5 Jahre und etwa 15 Entwickler um die gefundenen Schwierigkeiten zu beseitigen. Schon 2008 kann Watson hin und wieder ehemalige "Champions" schlagen, seit 2010 tut er dies regelmäßig.

2011 gewinnt Watson in einer echten Jeopardy Show gegen zwei der Top ehemalige Champions.

Danach wird es still um Watson, doch die Anwendungen in der Industrie und Wirtschaft laufen auf Hochtouren.

0.2 Wie funktioniert Watson

0.2.1 Was ist Jeopardy

Jeopardy ist eine Amerikanische Quizshow, bei der 3 Kandidaten gegeneinander antreten, darunter ein "Titelverteidiger". Der Gewinner der letzten Show. Die "Fragen" sind als Aussagen, Hinweise genannt, formuliert, die Antwort muss als Frage ausgedrückt werden.

z.B. Hinweis: "The answer to life the universe and everything"

Antwort: "What is 42?"

Die Hinweise kommen aus den verschiedensten Kategorien welche häufig witzige oder irreführende Namen haben. z.B. "Pink" "Words that rhyme with dog" "Presidents" "Washington"

Eine Spiel besteht aus 3 Runden: "Jeopardy!", "Double Jeopardy!", und "Final Jeopardy!"

1. Runde: Jeopardy!

Es gibt 6 Kategorien mit jeweils 5 Fragen, die in aufsteigender Schwierigkeit und mit einem aufsteigenden Preis versehen angeordnet sind. Der wiederkehrende Champion darf beginnen und wählt einen Hinweis aus, welcher vom Host vorgelesen wird.

Nach einem Lichtsignal können die Spieler einen Buzzer betätigen. Der schnellste darf seine Antwort geben. Ist die Antwort falsch haben die anderen 2 Mitspieler eine weitere Chance zu "buzzen".

Die Spielstände werden durch den Preis der beantworteten Fragen berechnet. Bei einer richtigen Antwort wird der Preis der Frage auf den Spielstand addiert, bei einer falschen subtrahiert.

Im Falle, dass niemand die richtige Antwort gibt, wird diese vorgelesen und das Spiel geht mit dem nächsten Hinweis weiter.

Hinter einem Hinweis auf der Tafel verbirgt sich ein "Daily Double". Derjenige der den "Daily Double" entdeckt darf als einziger diesen Hinweis beantworten, doch als erstes muss er um einen Teil seines Spielstandes (Minimum 5\$ Maximum aktueller Spielstand) wetten.

Egal ob die Antwort richtig oder falsch war, darf der Kandidat den nächsten Hinweis auswählen.

2. Runde: Double Jeopardy!

Die 2. Runde verläuft gleich der ersten, nur dass die Hinweise doppelten Wert haben. Auch gibt es in dieser Runde 2 Daily Doubles.

Der Kandidat mit dem niedrigsten Spielstand darf beginnen. Kandidaten die diese Runde mit einem Spielstand von 0\$ oder niedriger verlassen, werden eliminiert.

3. Runde: Final Jeopardy!

In der letzten Runde gibt es nur einen einzigen Hinweis.

Die Kategorie wird vor Beginn der Runde bekannt gegeben und es werden Barrikaden zwischen den verbleibenden Kandidaten errichtet. Jeder gibt schriftlich eine Wette ab zwischen 0\$ und dem aktuellen Spielstand, dann wird der Hinweis vorgelesen und die Spieler haben 30 Sekunden Zeit ihn schriftlich zu beantworten. Die Antworten werden in Reihenfolge der Spielstände veröffentlicht. Wieder wird bei richtigen Antworten die Wette addiert, bei falschen subtrahiert.

Der Kandidat mit dem höchsten Spielstand gewinnt

Die Schwierigkeiten von Jeopardy

Die Schwierigkeiten (aus menschlicher und maschineller Sicht) sind folgende:

1. Die Fragen kommen aus einem weiten Feld von Kategorien mit komplizierten, in natürlicher Sprache gehaltenen Hinweisen.
2. Es wird eine hohe Präzision und eine Abschätzung zum Vertrauen in die eigene Antwort verlangt.
3. Es muss schnell geantwortet werden. Insbesondere schneller als die Mitspieler

Für Watson sind diese Probleme natürlich anders, als für Menschen.

Das größte Problem für Watson ist die natürliche Sprache. Er muss erkennen ob die Kategorie, aus der der Hinweis stammt, ihm helfen kann oder nicht, worauf sich die Frage eigentlich bezieht und natürlich in seinen Quellen (welche großteils auch in natürlicher Sprache verfasst sind) eine Antwort finden.

Als nächstes brauchen wir eine hohe Präzision, Geschwindigkeit und ein sehr gutes Abschätzungsvermögen ob die gefundene Antwort richtig sein könnte. Nach analysieren von vielen Jeopardy Spielen haben die Entwickler ermittelt, dass die Champions mindestens 70% der Zeit Buzzen (also genug Vertrauen in ihre Antwort haben) und mindestens 80% dieser der Antworten richtig sein müssen, und das obwohl wir nur wenige Sekunden haben um eine Antwort zu finden und schneller als alle anderen zu sein.

Das Finden der Antwort geht los, sobald der Hinweis aufgedeckt wurde, das Buzzen erst, wenn ein Signal ertönt (nachdem der Hinweis zu Ende vorgelesen wurde). Wenn Watson zu diesem Zeitpunkt eine Antwort hat (die ihm sicher genug erscheint) dann könnte man meinen er sei natürlich schneller am Buzzer als jeder Mensch.

Allerdings erwarten manche Kandidaten das Licht. Etwas was Watson nicht kann. So verliert sogar der Computer das Reaktionsspiel.

0.2.2 DeepQA

Watson vs DeepQA

Watson ist eigentlich nur das System um Jeopardy! zu spielen. Das öffentliche "Gesicht" der Technologie. Die zugrundeliegende Architektur nennt sich DeepQA.

Question-Answering Systeme haben zwar eine lange Geschichte, aber vor Watson waren sie nicht entwickelt genug um signifikante Anwendungen zu finden.

Hardware

Watson ist tatsächlich kein Supercomputer. Seine Linpack Performance steht bei ca. 80 TeraFLOPs top was kaum der Hälfte der Leistung entspricht die gebraucht wird um es auf die Top 500 Supercomputer Liste zu schaffen.

Der Rechenverbund besteht aus 90 Power 750 Servern mit insgesamt 16 TB RAM auf der auch sämtliche Quellen sind. Watson benutzt ausschließlich Arbeitsspeicher um wertvolle Zeit zu sparen wenn er eine Antwort finden soll.

Jeder Server besitzt einen mit 3,5 GHz getakteten Power7 8-Kern Prozessor, wobei jeder Kern bis zu 4 Threads gleichzeitig ausführt.

Einfacher ausgedrückt: er kann 500GB in einer Sekunde verarbeiten.

Noch einfacher: dies entspricht etwa ca. 1 000 000 Bücher pro Sekunde zu lesen.

Warum ist DeepQA besonders?

DeepQA stützt sich zwar größtenteils auf bereits bekannte Technologien, aber keine davon alleine ist für den Erfolg verantwortlich. Erst das Zusammenspiel aller macht DeepQA so erfolgreich.

3 Faktoren spielen allerdings eine wichtige Rolle:

1. das System operiert rund um eine Pipeline in der Hypothesen sequenziell überprüft werden
2. die Pipeline-Schnittstellen sind normiert so das viele Ansätze parallel verfolgt werden können. So steht das Team sich nicht selbst im Weg und kann gleichzeitig an vielen Baustellen arbeiten
3. es werden Gewichtsfunktionen verwendet um partielle Evidenz-Werte zu einem gemeinsamen Score mit Hilfe von gelernten Gewichtungen zu verwandeln.

Der tatsächliche Vorgang in DeepQA, zum Beispiel das lösen eines Jeopardy Hinweises durch Watson, erfolgt in mehreren Schritten.

Schritt 1: Fragenanalyse

Als erstes wird die Jeopardy Aussage analysiert um zu erkennen was gesucht wird. Ein Präprozessor und Syntaxanalyser bricht die Aussage in Elementarsätze herunter. Zum Beispiel eine Aussage wie "this person" oder "that year" führt zu einer linguistischen Kategorie in der dann die Antwort gesucht wird.

Die Entwickler verwenden lexikalische Kategorien und kamen beim studieren alter Jeopardy! Hinweise auf 2.500 Arten solcher Kategorien. Allerdings decken bereits 40 davon schon einen Großteil der Fragen ab.

In 11% der Fälle lies sich keine Kategorie finden und in diesem Fall verzichtet der Computer ganz auf eine Antwort.

Schritt 2: Kandidatengenerierung

Nach der Analyse von Frage und Kategorie, können verschiedene Methoden verwendet werden um Hypothesen zu generieren. Diese wiederum erzeugen hunderte Kandidaten und nach einer ersten Überprüfung bleiben etwa 100 übrig.

In Watson sind etwa 100 verschiedene Methoden implementiert. Zum Beispiel:

1. Stichworte in Wikipedia suchen und den Titel der Seite als Antwort verwenden
2. in strukturierten Datenbanken suchen (z.B DBpedia)
3. in Listen der Kategorie suchen (z.B Liste Amerikanischer Städte)

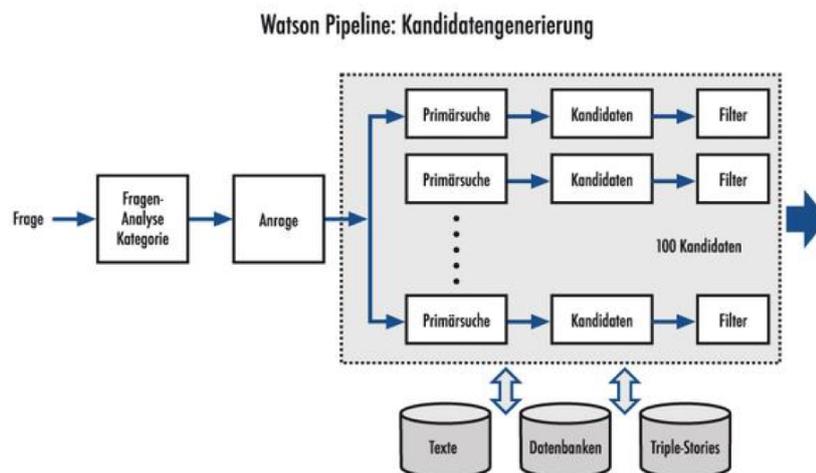


Abbildung 1: Kandidatengenerierung

Schritt 3: Kandidatenbewertung

Im letzten Teil der Pipeline werden die Kandidaten weiter ausgedünnt und auch hier arbeiten wieder viele Methoden parallel.

zum Beispiel werden Kandidaten wieder zurück in andere Quellen gegeben. Falls der Hinweis oder Teile wiederum darin vorkommen steigt das Vertrauen in die Hypothese.

Auch Zeit und Raum werden überprüft. Passen diese nicht zum Hinweis, wird aussortiert

Bei der Überprüfung von 100 Hypothesen werden bis zu 100.000 Evidenzobjekte zur Hilfe gerufen.

Am Ende bleiben nur wenige Kandidaten übrig die zur finalen Überprüfung und dann durch ein Ranking geschickt werden. Zu den verbleibenden Hypothesen gibt es dann noch jeweils mehrere hundert Scores, die die Wichtigkeit und das Vertrauen in das Ergebnis abschätzen. Diese numerischen Noten werden einfach linear kombiniert und der berechnete gewichtete Mittelwert wird dann zwischen 0 und 1 skaliert. (Stichwort logistische Regression) Die Gewichte werden anhand einer Datenbank maschinell gelernt. So werden am Jeopardy Beispiel z.B Gewichte für Wortspiele oder geographische Nähe vergeben. Danach gibt Watson die verbleibenden Kandidaten zusammen mit dem Numerischen Score aus.

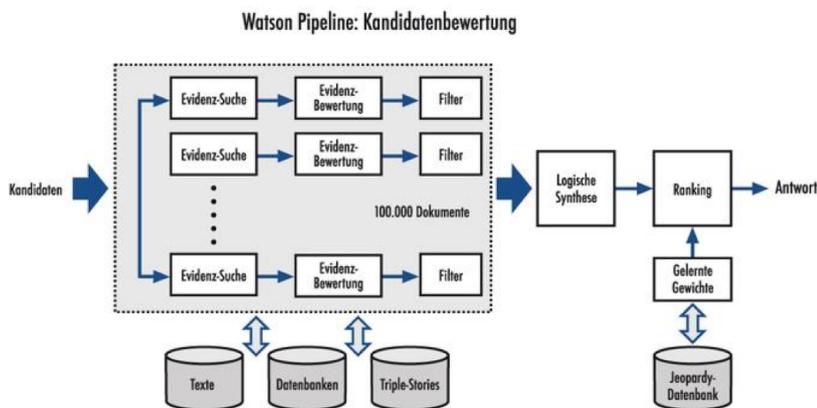


Abbildung 2: Kandidatenbewertung

0.2.3 Schwächen

Komplexe logische Überlegungen kann Watson nicht durchführen. Weder Beweise noch Schlussfolgerungen sind für das System zu bewältigen, denn außer ein paar wenigen transitiven Beziehungen und Synonymen kann er nicht "weiter denken."

Dies ist auch der nächste Schritt der Entwickler: Inferenzmechanismen in DeepQA einbauen

Echte Sprache, Schwere Sprache

Natürliche Sprache ist simpel und intuitiv, zumindest für Menschen. Die Zweideutigkeit von Ausdrücken ist eher lustig als frustrierend.

"One morning I shot an elephant in my pajamas. How he got in my
pajamas, I don't know."

- Groucho Mark, Animal Crackers

Warum ist das witzig? Unser Gehirn denkt gar nicht daran die Möglichkeit in Betracht zu ziehen, dass der Elefant im Schlafanzug war. Aber warum? Intuitiv wissen wir die relative Größe der involvierten Objekte. Zusätzlich wissen wir, wie viel Wahrscheinlicher es ist, das ein Mensch einen Schlafanzug trägt im Vergleich zu einem Elefanten. Aber wenn wir diese Information nicht hätten, dann wären beide Optionen gleich wahrscheinlich.

Wenn wir Sprache verstehen schätzen wir intuitiv und logisch alles anhand unseres Wissens ab. Wir benutzen Kontext um Dinge verständlich zu machen die eigentlich mehrdeutig oder unpräzise formuliert sind. Ist also die Lösung, Computer natürliche Sprache verstehen zu lassen, ihnen alles Wissen der Welt zu geben?

Die kurze Antwort ist: Jein

In kleinen Bereichen geht das gut, Watson z.B nutzt diesen Ansatz (unter anderem) um mögliche Antworten abzuschätzen. Eine Frage wie "Die nördlichste Hauptstadt" mit Auswahl Manila, Kathmandu und Jakarta kann man einfach nachprüfen in dem man präzise geographische Daten vergleicht. (Kathmandu ist die Antwort). Aber was ist mit:

"If you're standing, it's the direction you should look to check out the
wainscoting."

Welches Wissen der Welt musst du haben um hier eine Antwort zu geben? Das alleinige Wissen was "wainscoting" ist, reicht hier nicht. Die relative Größe zum Menschen ist entscheidend. Aber woher soll die Maschine wissen, dass sie Relationen abschätzen muss, wenn kein Stichwort für Relationen überhaupt in der Frage vorkommt?

Dies ist ein Beispiel für einen Hinweis, den Watson nicht beantworten kann.

Allerdings können wir das auch andersherum.

$$\frac{(\ln(12345678 \cdot \pi))^2}{34567.46}$$

Ist dieser Ausdruck größer oder kleiner 1?

Ein Mensch kann dies nicht direkt sagen, der Computer kann uns sofort ausgeben, dass das Ergebnis 0.00885 ist. Sobald die Maschine weiß, was wir von ihr wollen, ist also ein Großteil des Problems geschafft. Sobald aber Frage oder Antwort in natürlicher Sprache gehalten sind, gelten die Stärken des Computers nicht mehr.

In der Jeopardy! Challenge sind nicht nur Frage und Antwort in natürlicher Sprache sondern auch die meisten Quellen.

Betrachten wir die Frage "Wo wurde Einstein geboren?"

Haben wir einfach eine Tabelle mit Geburtsorten bedeutender Menschen, gehen wir einfach zur Zeile Einstein und gucken in der passenden Spalte nach.

Aber was wenn die Antwort eingebettet ist?

"One day, from among his city views of Ulm, Otto chose a water color to send to Albert Einstein as a remembrance of Einstein's birthplace."

Die Antwort ist da. Aber wo? Warum genau dieser Satz? Wir müssen eine Menge Sätze lesen bevor wir entscheiden können dass dieser relevant sein könnte, und selbst wenn wir die Passage bereits haben ist es schwierig die Antwort zu extrahieren. Wir müssen alle Menschen und Orte und ihre Verbindung zu Einstein aufdröseln um die Antwort zu finden. Natürlich gibt es auch einfachere Sätze, aber solange Computer nicht in der Lage sind solche Sätze zu verstehen ist der Weg zur "True AI" noch ein sehr weiter.

0.3 Anwendungen

0.3.1 Generell

Deep QA ist vor allem ein Software-Rahmenwerk. Man kann dann verschiedene Verfahren und Methoden aus der Textretrieval anbinden (Stichwort Mustererkennung: Ensemble Methoden) Die Schnittstellen in der Pipeline sind standardisiert, das heißt neue Generierungsverfahren können direkt getestet werden indem man eine neue Zeile in die Pipeline Matrix man addiert einfach eine neuen Zeile ähnlich mit Gütekriterien man addiert neue Spalten (und den dazugehörigen Score) gegeben das Basissystem können also mit wenig aufwand geprüft werden ob eine Idee etwas zum Gesamtergebnis beiträgt z.B. speziell für Jeopardy! ein Wortspiel-Kriterium eingeführt, das in vielen Fällen sehr hilfreich war.

0.3.2 die Firma IBM Watson

Im Januar 2014 gab IBM bekannt, eine Business Unit um Watson zu eröffnen die von Senior Vice President Michael Rhodin geleitet wird. Das HQ der Watson Group liegt in New York's Silicon Alley und beschäftigt ca 2 000 Menschen. IBM investierte ca 1 Milliarde (1 000 000 000) Dollar und die Division zu starten.

Die Entwicklung fokussiert sich auf 3 Cloud-Basierte Services:

1. Watson Discovery Advisor für Research und Development in der Pharma und Bio-Technologie Industrie
2. Watson Engagement Advisor für Self-Service Applikationen die Fragen in natürliche Sprache von Nutzern beantworten
3. Watson Explorer der Firmen helfen soll share-data basierte Einblicke durch föderierte Suche zu vereinfachen

Außerdem gibt es einen Hundert Millionen (100 000 000) Dollar Fund um Entwicklung für "Kognitive" Applikationen voranzutreiben Laut IBM hat der Cloud-Delivered Enterprise-Ready Watson seine Geschwindigkeit (seit 2011) um das 24 fache verbessert. das sind 2,300% Verbesserung in der Performance. Seine Größe verringerte sich dabei um etwa 90% von einem großen Raum zu etwa der Größe von 3 gestapelten Pizzakartons IBM CEO Virginia Rometty sagte, sie möchte, dass Watson innerhalb der nächsten 10 Jahre ein jährliches Einkommen von 10 Milliarden (10 000 000 000) Dollar generiert.

0.3.3 Medizin

Als eine neue Aufgabe sind wie erwähnt medizinische Systeme ausgewählt worden. Es gibt mittlerweile erfolgreiche Programme, die den Ärzten zu ge-

gegebenen Symptomen, Laborwerten, Diagnostik und Anamnese Therapie und Medikamentenvorschläge gibt, zusammen mit Beurteilungen welche Wechselwirkungen zu erwarten sind oder welcher Therapieansatz aufgrund von Laborwerten etc. nicht zu empfehlen ist.

0.3.4 Olli derBus

Auf der CeBit 2017 Wurde Olli vorgestellt. Ein Kleinbus der autonom fährt. Watson soll sich per Sprachsoftware mit den Fahrgästen unterhalten und zum Beispiel Restaurant-Tipps geben. In Washington, Las Vegas und Miami wird er bereits getestet. In Deutschland gibt es ihn bisher allerdings nur auf einem Privatgelände in Berlin.

0.3.5 Onlinepräsenz

Auf verschiedenen Websites kann man selbst mit Watson arbeiten:

Auf <http://www.donotpay.co.uk/> kann man sich von Watson Knöllchen anfechten lassen oder eine offizielle Erklärung zum HIV Status abgeben. Auf <https://personality-insights-livedemo.mybluemix.net/> kann man seinen oder fremde Twitter-Accounts, Reden, oder andere Texte nach Stimmung und Charakter analysieren lassen. Und die IBM Developer Cloud bietet kostenpflichtige und kostenfreie Optionen an, seinen eigenen Chatbot zu Programmieren oder Watson für das eigene Unternehmen zu nutzen.

0.3.6 Weitere

1. Chatterbot: Watson wird als Konversationstool für Kinderspielzeuge verwendet
2. Teaching Assistant: ein Professor an der Georgia Tech erschuf einen virtuellen TA namens "Jill" der Fragen von Studenten beantwortete sobald er sich 97% sicher mit der Antwort war.
3. Wetter: Im August 2016 verriet IBM dass es Watson zur Wettervorhersage verwendet würde. Genauer: er würde Daten von über 200 000 Offiziellen und Persönlichen Stationen analysieren zusammen mit anderen Daten von anderen Quellen unter dem Projektnamen Deep Thunder
4. Steuerberatung: seit dem 5.2.2017 gibt es ein Watson-basiertes Programm der Firma H&R Block zum ausfüllen deiner (Amerikanischen) Steuererklärung

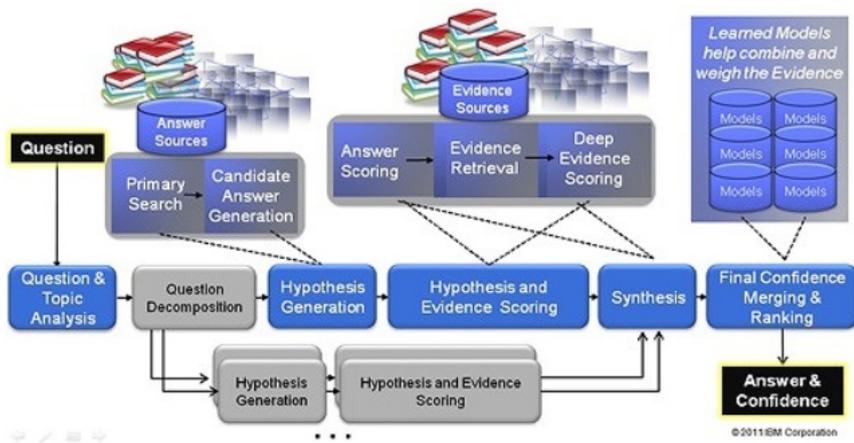


Abbildung 3: Pipeline Überblick

Literaturverzeichnis

- [1] *Building Watson: An Overview of the DeepQA Project*
AI Magazine, Herbst 2010
- [2] *Warum Watson ein Durchbruch ist*
Raúl Rojas, 24.3.2012, heide.de/telepolis
- [3] *Unsupervised Entity-Relation Analysis in IBM Watson*
Aditya Kalyanpur, J. William Murdock, ACS-2015 (Article 12)
- [4] *Multiparticipant chat analysis: A survey*
David C. Uthus, David W. Aha, 2013, elsevier.com
- [5] *IBM Watson Research Website*
http://researcher.watson.ibm.com/researcher/view_group_subpage.php?id=2162
- [6] *IBM Developer Cloud*
<https://www.ibm.com/watson/developercloud/services-catalog.html>
- [7] *Wikipedia* [https://en.wikipedia.org/wiki/Watson_\(computer\)](https://en.wikipedia.org/wiki/Watson_(computer))