

Daniel Spaude

Watson und DeepQA

Überblick

Aspekte des Vortrags

- Watson: Einführung und kurzer historischer Rückblick
- technischer Überblick und Bestandteile von Watson

Anliegen dieses Vortrags

1. Einführung in Problemstellung und Motivation hinter Watson / DeepQA
2. Orientierungswissen für den weiteren Seminarverlauf und die vorgestellten Artikel

1. Watson: Einführung und historischer Rückblick

Watson is about the very challenging AI-problem of:

**open-domain question-answering
(QA)**

1. Watson: Einführung und historischer Rückblick

open-domain question-answering (QA)

berührt Gebiete wie:

- information retrieval (IR)
- natural language processing (NLP)
- knowledge representation and reasoning
- machine learning
- human-computer interfaces (HCI)

Open domain, weil...

es sich nicht auf ein vorher abgestecktes Gebiet beschränkt

stattdessen: gesamtes, durch Sprache ausdrückbares Lebensumfeld des Menschen

Watson wants to play...

Jeopardy!

Jeopardy!

Warum Jeopardy?

Quiz-Format, welches besonders die Eigenheiten natürlicher menschlicher Sprache integriert

seit Jahrzehnten laufendes und international bekanntes US-Fernsehformat
--> Publicity für IBM

Jeopardy!

Wikipedia:

"Unique answer-and-question format in which contestants are presented with clues in the form of answers, and must phrase their responses in question form."

("Einzigartiges Frage-Antwort-Format, in welchem die Kandidaten Hinweise in Form von Antworten präsentiert bekommen und darauf mit der passenden Frage antworten müssen")

Jeopardy-Fragen: Beispiele

The Father of Our Country; he didn't really chop down a cherry tree.

Jeopardy-Fragen: Beispiele

Who is George Washington?

Jeopardy-Fragen: Beispiele

The Rolling Stones formed in 1962, when this lead vocalist was a lad of 18.

Jeopardy-Fragen: Beispiele

Mick Jagger

Jeopardy!

Um in Jeopardy zu gewinnen, muss man...

- eine Menge wissen
- belesen sein
- Sprache und ihre Nuancen verstehen
- schnell reagieren

Watson wants to win against...

the world champions

Watson wants to win against...

Ken Jennings

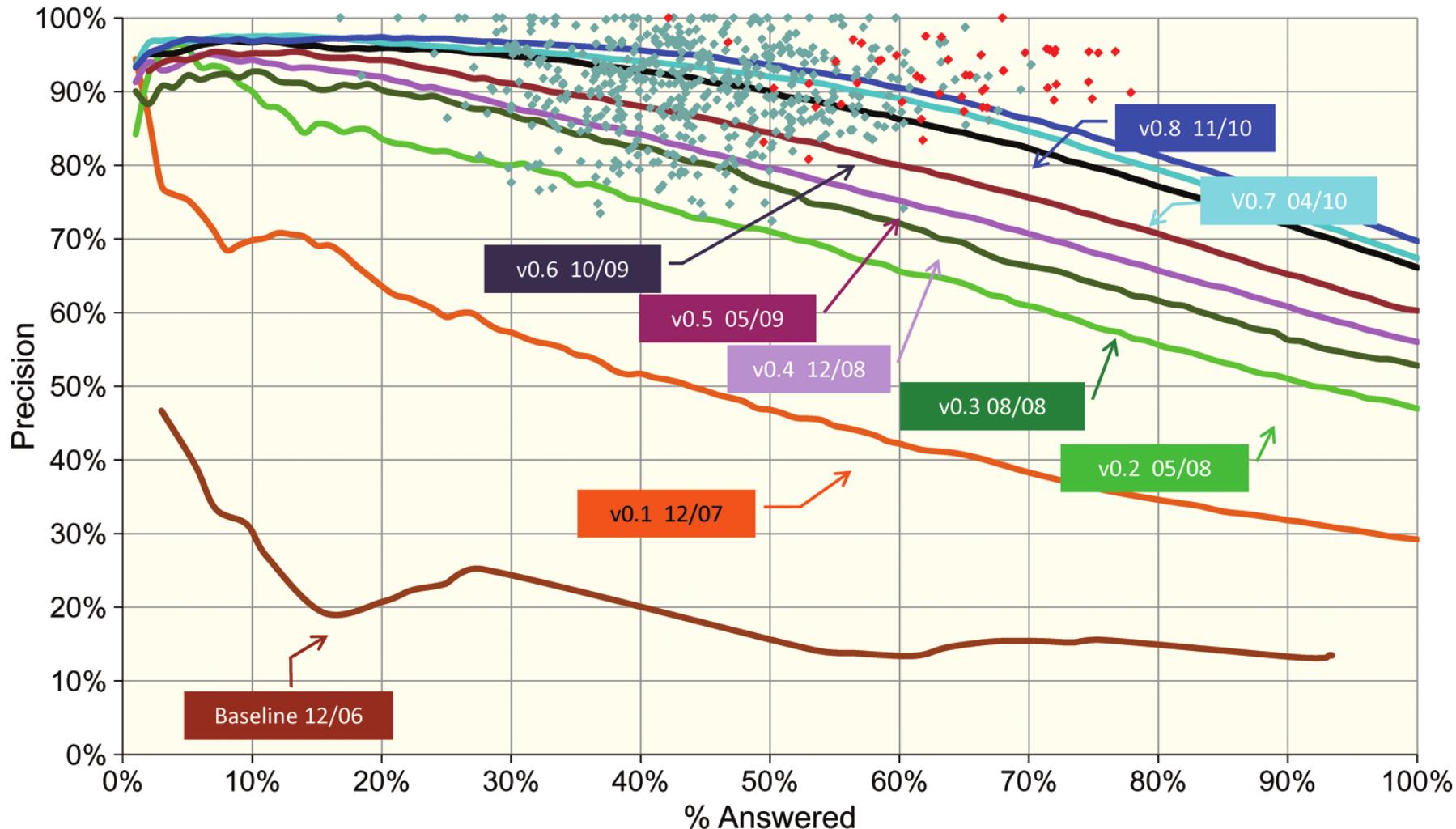
Gewann hintereinander 72 Partien im Jahr 2004 und gewann insgesamt \$2,520,700

Brach damit mehrere Rekorde sowohl in Jeopardy! als auch amerikanischen Gameshows allgemein.



The winners cloud

Watsons goal: 85% Precision@70



Unstructured information

Der größte Anteil menschlicher Kommunikation ist unstrukturierter Natur

- Bücher
- Journale
- Technische Handbücher
- Biographien
- Blogs, ...

Die Bedeutungen sind oft implizit und kontextabhängig und müssen anhand von Hintergrundinformationen abgeleitet werden.

Unstructured information

- natürlicher Umgang für den Menschen
- aber schwierig für Computer,
Bedeutungen abzuleiten
- strukturierte Daten:
definiertes Datenschema, beispielsweise
relationales Datenbankmodell bei einer
Buchhaltungssoftware
oder auch, begrenzt: Facebook

Könnten Computer...

... Bedeutungen in unstrukturierten Inhalten erkennen und mit diesen arbeiten, ergäben sich daraus neue Anwendungsmöglichkeiten in:

Forschung, Medizin, Jura, ...

Bildung,

Dialogsysteme,

decision making in Wirtschaft und Politik, ...

POSTAGE
STAMP
HERE

GOOGLE

CLASSIC

QUERY:

IMAGES

NEWS

VIDEO

MAPS

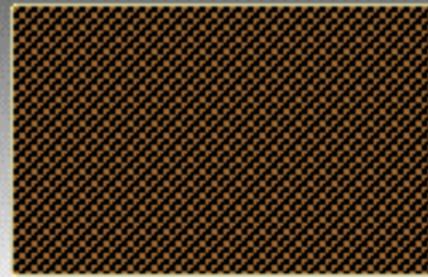
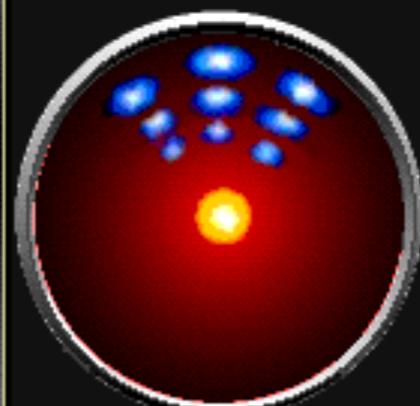
OTHER

SEND YOUR QUERY TO: GOOGLE INC., 1600 AMPHITHEATRE PARKWAY, MOUNTAIN VIEW, CA 94035, UNITED STATES

PLEASE ALLOW 30 DAYS FOR SEARCH RESULTS

VS

HAL 9000



[https://www.youtube.com/watch?
v=MA1hD3XRIh0](https://www.youtube.com/watch?v=MA1hD3XRIh0)



NLP Techniken

...leiten die Bedeutung eines Ausdrucks oder einer Phrase ab mittels Analyse von Syntax, Kontext und Gebrauchsmustern

NLP Techniken

Menschliche Sprache ist jedoch so komplex, variabel und mehrdeutig, dass dies eine bis heute nicht zufriedenstellend gelöste Herausforderung darstellt.

NLP Techniken

Jahrzehnte an Forschung haben mehrere verschiedene spezialisierte Techniken hervorgebracht.

Jede Technik operiert auf der Sprache auf unterschiedlichem Level und unterschiedlichen Aspekten des Verstehensprozesses.

NLP Techniken

Keine dieser Techniken ist perfekt oder komplett im Entschlüsseln der intendierten Bedeutung.

Denn menschliche Sprache ist im Gegensatz zu einer Programmiersprache oder auch Schach kein mathematisch formalisierbares Konstrukt.

<https://www.youtube.com/watch?v=6ay17a7mElk>



West Virginia

Beispiel fuer Mehrdeutigkeit von Sprache

Finde die exakte Definition des Wortes "Spiel"!

Ludwig Wittgenstein:

"Die Bedeutung eines Wortes ergibt sich aus seiner Verwendung in der Sprache."

Sprachspiel



Die verschiedenen NLP Techniken

...werden und sollten sich auch weiterhin unabhängig voneinander verbessern

Die verschiedenen NLP Techniken

Noch wichtiger jedoch: Integration dieser unterschiedlichen Techniken und Algorithmen, um ein verbessertes ganzheitliches Sprachverständnis des Systems zu erzielen

Integration der unterschiedlichen NLP-Techniken...

...ist eines von Watson's Hauptleistungen

Apache UIMA

Unstructured Information Management
Architecture

Ebenfalls von IBM entwickelt und 2006 zur
Apache Foundation übergeben

-> Open Source

Apache UIMA

Software-Architektur, die eine flexible Plattform bereitstellt

...zur Integration diverser Sammlungen von Text, Sprache und Bilddaten

...und unabhängig ist vom verwendeten Algorithmus, der Programmiersprache oder dem zugrunde liegenden Domänenmodell.

Apache UIMA

Software-Architektur, welche die essentielle Infrastruktur bereitstellt...

...um mehrere hunderte verschiedener, kooperierender Algorithmen zusammenzubringen.

Jeder dieser Algorithmen führt für sich allein genommen relativ einfache language processing Aufgaben durch.

Apache UIMA

Jeder dieser Algorithmen führt für sich allein genommen relativ einfache language processing Aufgaben durch.

Marvin Minsky's "Society of Mind"

Apache UIMA

Marvin Minsky's "Society of Mind"

„Welcher magische Trick macht uns intelligent?

Der Trick ist, daß es keinen Trick gibt. Die Macht der Intelligenz stammt aus unserer ungeheuren Diversität und nicht aus einem einzelnen perfekten Prinzip.“

IBM's vorhergehende Expertise im Bereich open domain QA

PIQUANT:

QA System, Entwicklungsstart: 1999

Single Ontology

Combined Parsing

Information Extraction

Search

To generate answers to factoid-style questions.

IBM's Expertise im Bereich open domain QA

PIQUANT:

Performance von Piquant war deutlich zu schlecht für die Jeopardy-Teilnahme

Und eine Weiterentwicklung auf Grundlage von Piquant erwies sich als zu schwierig.

IBM's Expertise im Bereich open domain QA

PIQUANT:

- nutzte statische, vorher fest definierte Antworttypen
- also Klassen von Konzepten, auf welche eine Frage abzielt
- Beispiele: Menschen/Gesellschaft, Orte, Historische Daten, Zahlen, ...
- Der Umfang der möglichen Konzeptklassen in Jeopardy ist jedoch zu groß

"Radikaler" Ansatz

Annahme von Teamleiter Ferrucci:

gezielter, konzentrierter Versuch im großen Stil

bringt mehr als

jahrelange, fortlaufende inkrementelle Weiterentwicklung

Rahmenbedingungen und Herausforderung bzgl. Jeopardy

"Completely self-contained" - Keine Internetverbindung, keine Websuche, keinerlei Interaktion mit anderen Systemen oder Personen

--> gleiche Regeln wie bei menschlichen Teilnehmern

Rahmenbedingungen und Herausforderung bzgl. Jeopardy

Die Frage muss:

- in durchschnittlich 3 Sekunden aufgenommen und verstanden werden
- in Zusammenhang mit dem zuvor "Gelernten" gebracht werden

Und es muss:

- die beste Antwort bestimmt werden
- entschieden werden, ob der Buzzer gedrückt wird oder nicht

Rahmenbedingungen und Herausforderung bzgl. Jeopardy

- Google-ähnliche Keyword-Suche reicht nicht aus
 - welches Suchergebnis sollte gewählt werden?
 - sollte überhaupt eines gewählt werden?
 - welche Phrase eines Suchergebnisses oder welche Kombination aus Suchergebnissen stellt die korrekt formulierte Antwort da?

Inside DeepQA: The Architecture underlying Watson

Nachfolgend: Überblick über

- die Gesamtarchitektur
- die relevanten Systembestandteile
- und die entsprechenden Artikel

Inside DeepQA: The Architecture underlying Watson

Zwei wichtige technische Artefakte entstanden zu Beginn der Entwicklung von Watson:

1. DeepQA

- erweiterbare Softwarearchitektur für QA-Systeme

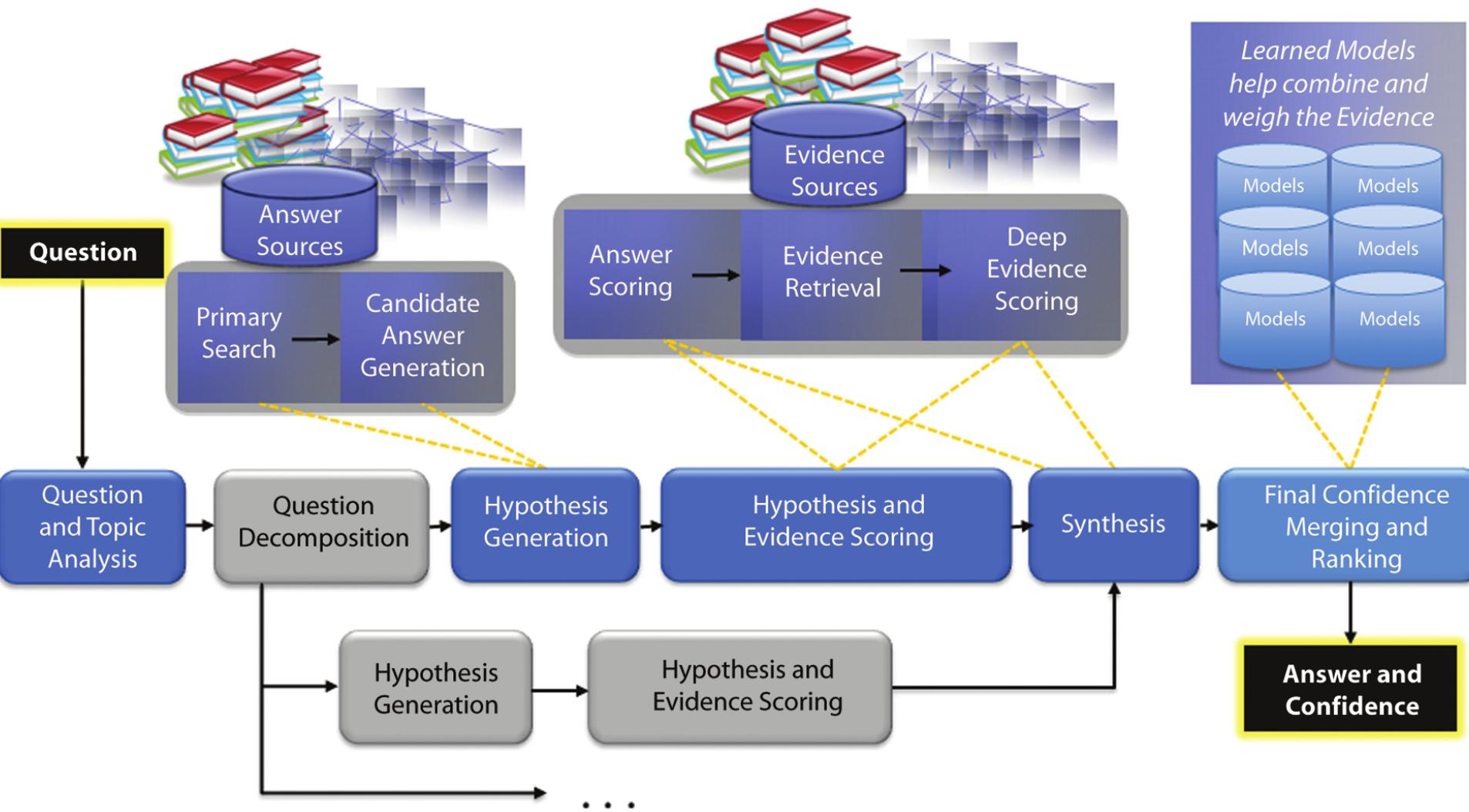
2. AdaptWatson

- Methodologie zur raschen Weiterentwicklung und Integration der Vielzahl an Kernalgorithmen

DeepQA Architektur

- Grundlegende Architektur von Watson
- mehrstufige processing pipeline
- Jede Stufe kann durch mehrere verschiedene Implementierungen abgedeckt werden
 - mit jeweils alternativem Ergebnis
- Massiv parallele Verarbeitung

DeepQA Architektur



DeepQA Architektur

- Annahme: keine Komponente versteht für sich genommen die Frage vollständig
- Stattdessen: mehrere potentielle Antworten (candidate answers) werden ermittelt
 - durch Suche in vielen unterschiedlichen Quellen
 - anhand verschiedener Interpretationen der Frage
- Wahl der finalen Antwort findet erst statt, nachdem ausreichend Beweise für jede candidate answer (jeden Pfad durch das System) vorliegen

DeepQA Architektur

- mehrere hunderte Algorithmen
 - analysieren Beweise über verschiedene Dimensionen, wie:
 - Typ, Zeit, Geographie, Popularität,
 - passage support
 - Glaubwürdigkeit der Quelle
 - semantischer Bezug
- Ergebnis: hunderte "scores"
 - beschreiben den Grad, zu welchem ein Beweis eine Antwort unterstützt

DeepQA Architektur

- die Menge scores müssen in einen einzelnen aggregiert werden
 - dieser beschreibt die Wahrscheinlichkeit, dass eine Antwort korrekt ist

DeepQA Architektur

- für die optimale Gewichtung der unterschiedlichen scores untereinander:
 - Training von statistischen machine learning Algorithmen

DeepQA Architektur

- Schlussendlich:
- Rangliste der candidate answers
 - jede mit endgültigem score / Wahrscheinlichkeit
- Wenn die beste candidate answer über einem bestimmten Schwellwert (threshold) liegt, möchte Watson antworten

DeepQA Architektur

- die DeepQA processing pipeline dient auch Bezugsbasis für die meisten der folgenden papers
 - Beschreibung viele der entwickelten Einzelalgorithem einzelner processing Stufen

AdaptWatson Methodologie

- Nach der initialen Implementierung von DeepQA:
- Entwicklung der AdaptWatson-Methodologie

AdaptWatson Methodologie

- teamorientierter Prozess und
- Menge von Tools für rasche Fortschritte der hunderten Einzelalgorithmen bezüglich:
 - Forschung
 - Entwicklung/Implementierung
 - Integration
 - und Evaluation

AdaptWatson Methodologie

- Beispiel:
- Auswertung von über 8000 unabhängigen Experimenten
- 10 bis 20 GB Output eines einzelnen Experiments

AdaptWatson Methodologie

- Tools zur Auswertung der Experimente wurden entwickelt
- Auf Grundlage der Analyseergebnisse:
 - Erzeugung neuer Ideen für Algorithmen

AdaptWatson Methodologie

- Zweiwöchige, vollständige system-builds
- Motivation: regelmäßig aktuelle Daten über die baseline performance von Watson
 - 200 eight-core servers
 - eigenes Core-Team von Forschern

DeepQA und AdaptWatson

- beide Artefakte bildeten die Grundlage für die weitere Entwicklung von Watson
- viele der nachfolgenden Papers befassen sich mit einzelnen Kernelgorithmen basierend auf der Verwendung von DeepQA und AdaptWatson

Daniel Spaude

Watson und DeepQA

Teil 2

**Überblick über die
technischen
Komponenten**

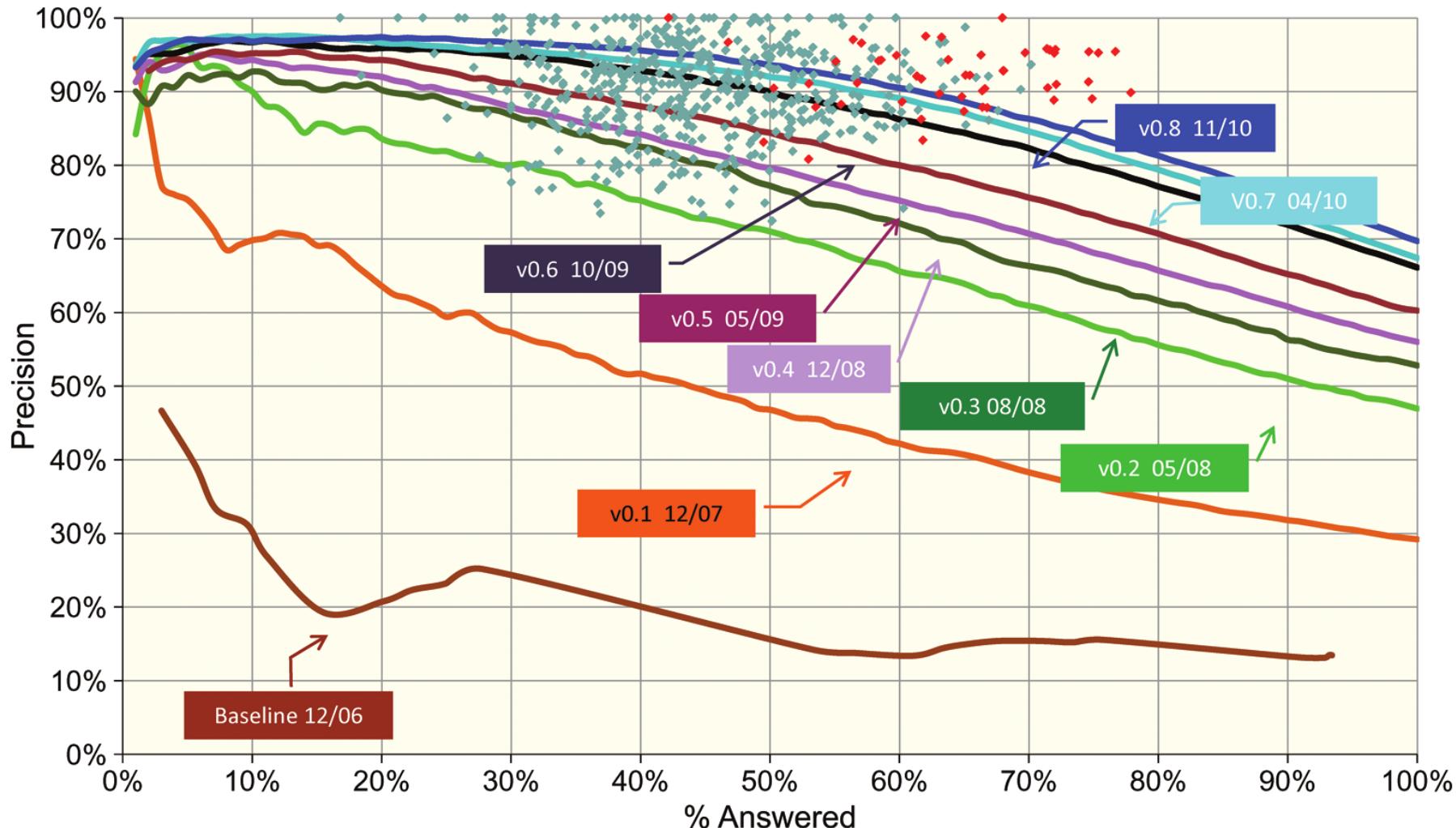
Kurzer Rückblick

Kurzer Rückblick

- open-domain question-answering (QA)
- Jeopardy
- Unstrukturierte Informationen / natürliche Sprache
 - Mehrdeutigkeiten, kontextabhängig, implizite Bedeutungen
 - leicht für den Menschen, schwer für einen Computer

The winners cloud

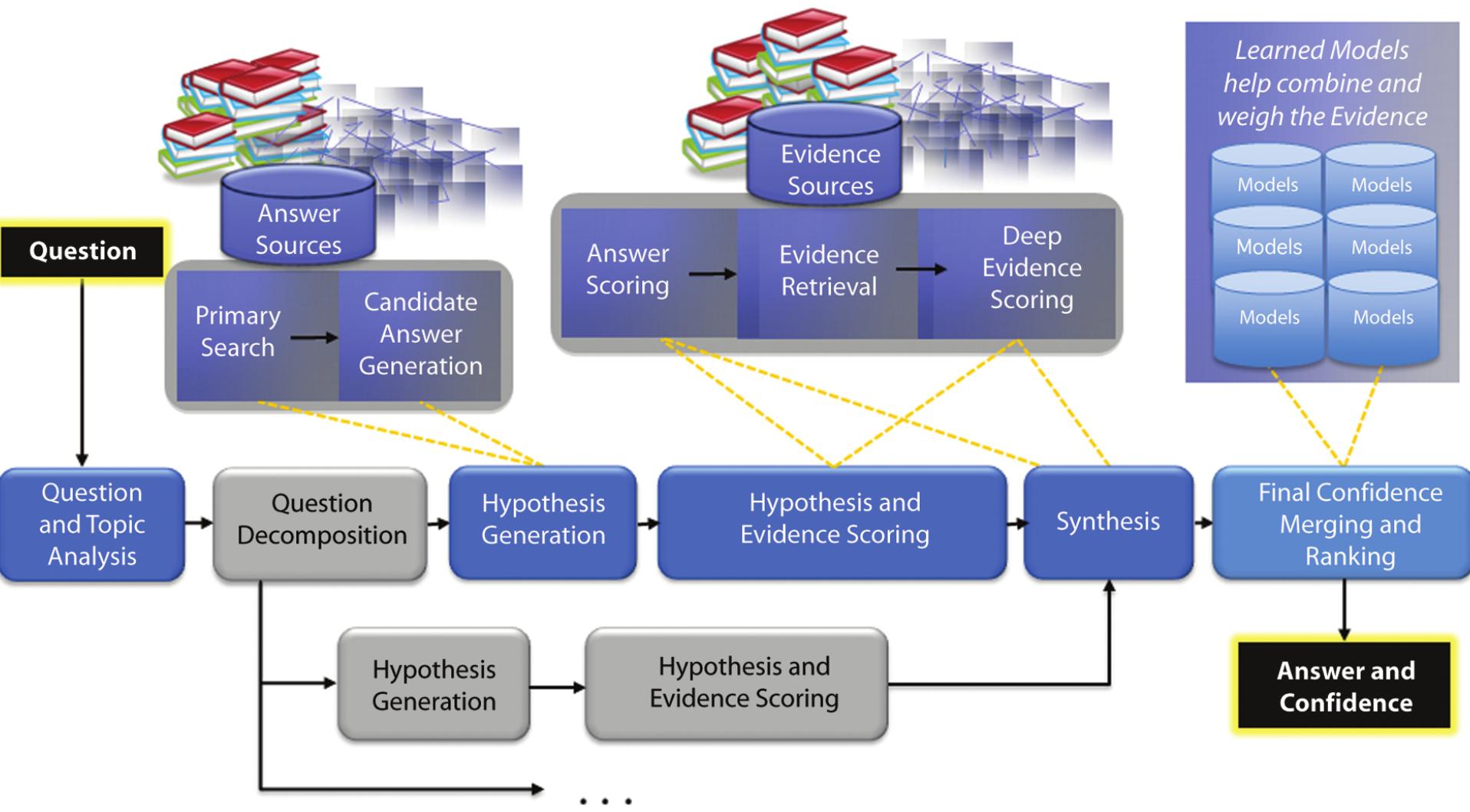
Watsons goal: 85% Precision@70



DeepQA Architektur

- Grundlegende Architektur von Watson
- mehrstufige processing pipeline
- Jede Stufe kann durch mehrere verschiedene Implementierungen abgedeckt werden
 - mit jeweils alternativem Ergebnis
- Massiv parallele Verarbeitung

DeepQA Architektur



DeepQA Architektur

- die DeepQA processing pipeline dient als Bezugsbasis für die meisten der folgenden papers

Los gehts

Die Fragen verstehen

- ... stellt große technische Herausforderung dar

Die Fragen verstehen

- lexical answer type (LAT)
 - Wort oder Phrase der Frage, welche auf die Klasse von Dingen hinweist, nach der gefragt wird

Die Fragen verstehen

- 1. Artikel "**Question Analysis: How Watson Reads a Clue**"
 - befasst sich mit dem Anfang der DeepQA Pipeline
 - beschreibt den Frage-Analyse-Algorithmus

Die Fragen verstehen

- 2. Artikel "**Deep Parsing Watson**"
 - stellt zwei Parsing-Komponenten vor:
 - ESG: English Slot Grammar parser
 - PAS: predicate-argument structure generator

Antwort- und Beweisquellen

- Watson ist komplett abgeschlossen
 - kein Internetzugriff
- daher: im Vorhinein bestimmen, welche Content-Quellen mit einbezogen werden

Antwort- und Beweisquellen

- 3. Artikel "**Textual Resource Acquisition and Engineering**"
- Content:
 - Auswahl
 - Auswertung
 - automatische Erweiterung

Antwort- und Beweisquellen

- 4. Artikel "**Automatic Knowledge Extraction from Documents**"
- Erstellung abgeleiteter Ressourcen
 - PRISMATIC

Bestimmung von potentiellen Antworten (candidate answers)

5. Artikel:

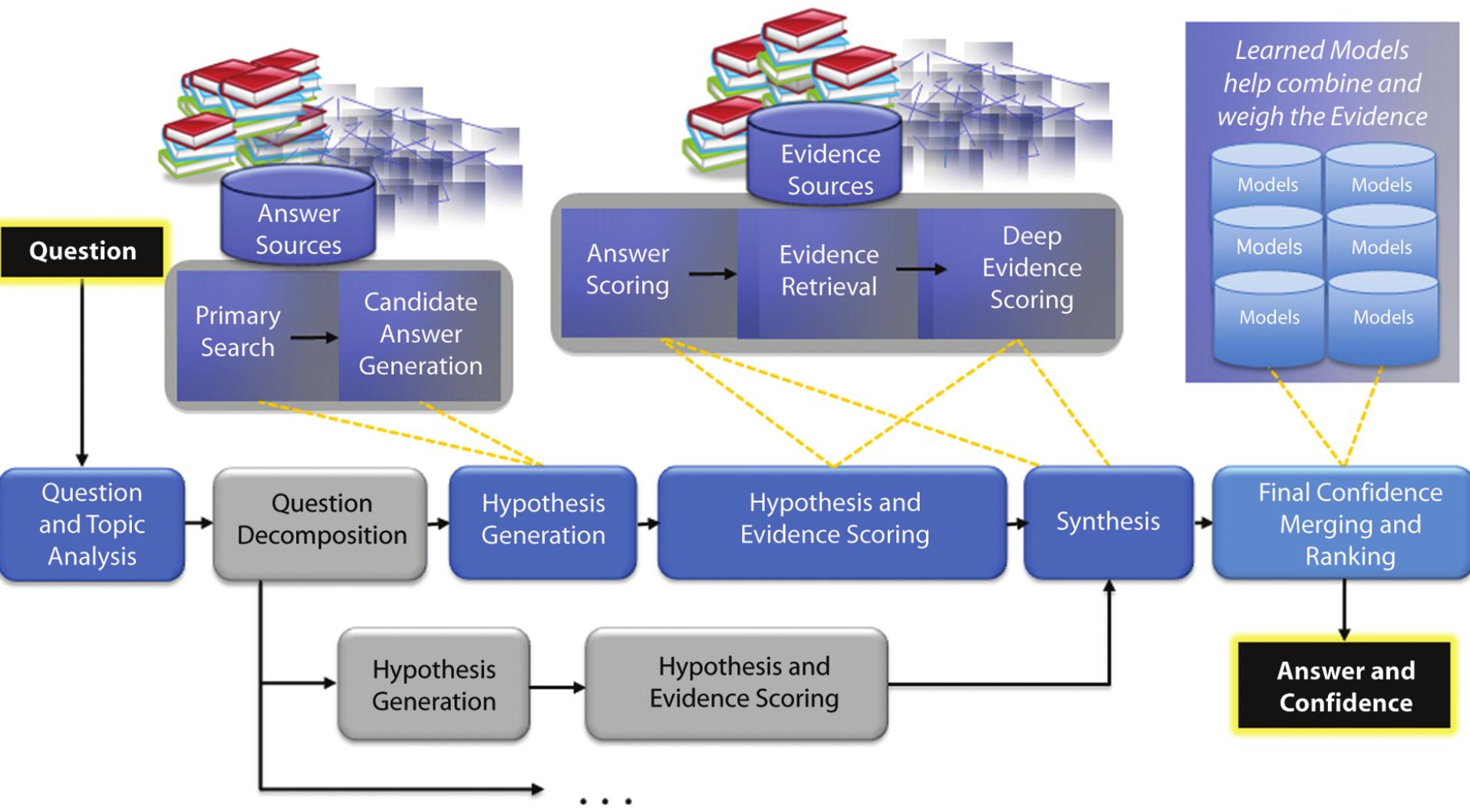
**"Finding Needles in the Haystack: Search
and Candidate Generation"**

- Komponenten zur
 - Suche
 - Kandidatenerzeugung
- candidate binary recall

Der Rest der DeepQA pipeline...

...widmet sich dem Finden der besten Antwort und dem Bestimmen einer akkuraten Wahrscheinlichkeit, dass diese korrekt ist.

DeepQA Architektur



Der richtige Antwort-Typ

...ist eine wichtige Klasse von Beweisen
(evidence), dass eine Antwort korrekt ist

6. Artikel:

"Typing Candidate Answers using Type Coercion"

Sammeln und Bewerten von Beweisen

Nach dem Erstellen mehrerer potentieller Antworten (candidate answers):

Sammeln und Bewerten (scoring) von zusätzlichen Beweisen

Hiermit befassen sich insgesamt drei Artikel.

Sammeln und Bewerten von Beweisen

Artikel 7

"Textual Evidence Gathering and Analysis"

**Eine Art von von Beweis:
passage evidence**

Sammeln und Bewerten von Beweisen

**Artikel 8 "Relation Extraction and Scoring in
DeepQA"**

Sammeln und Bewerten von Beweisen

**Artikel 9 "Structured Data and Inference in
DeepQA"**

Puzzles, Final Jeopardy! und andere spezielle Fragetypen

Artikel 10 "Special Questions and Techniques"

Missing links

Artikel 11 "Identifying implicit relationships"

Fragen hinunterbrechen

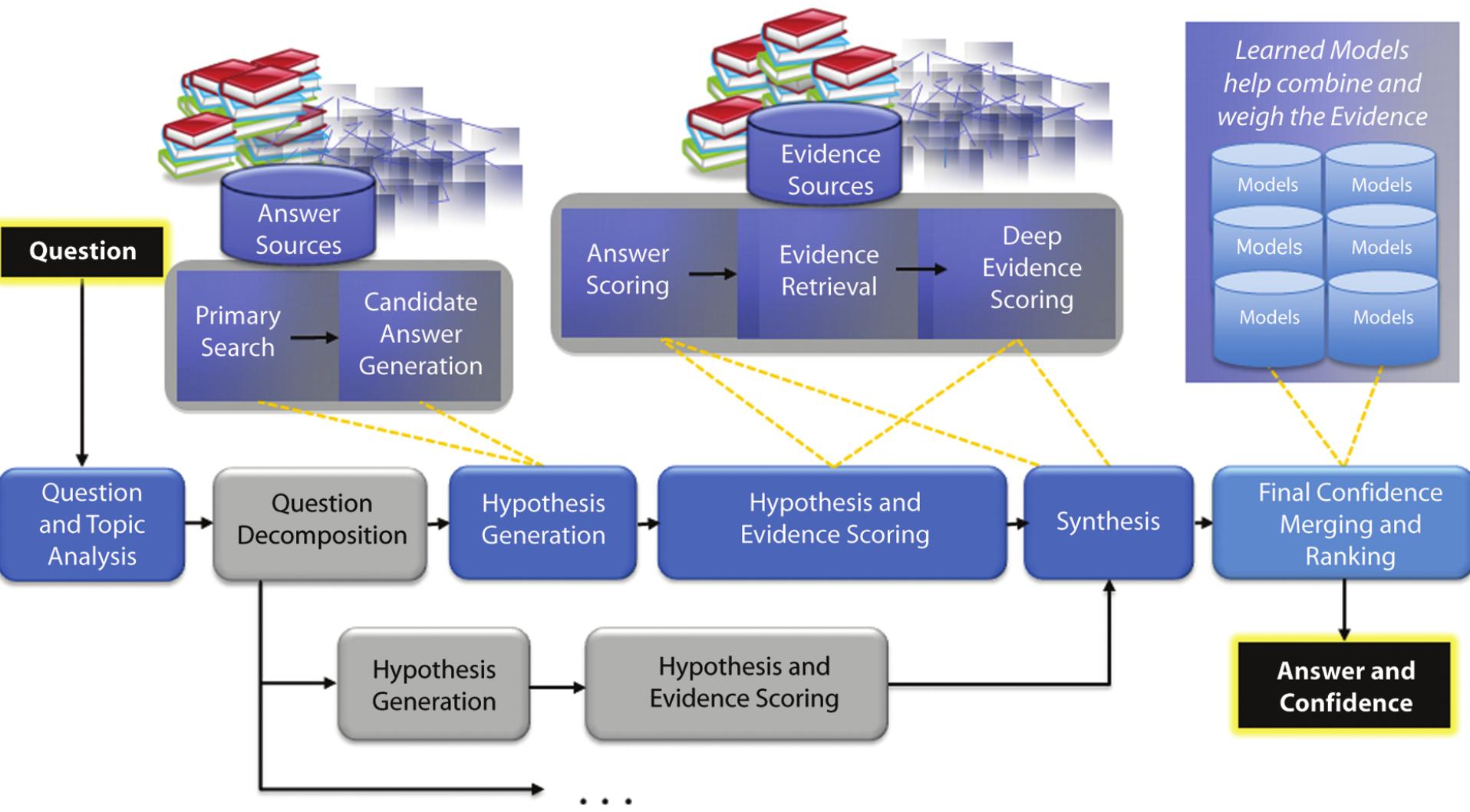
**Artikel 12 "Fact-based question
decomposition in DeepQA"**

Zusammenführen von Beweisen und Kombinieren von Überzeugungswerten

Artikel 13 "A framework for merging and ranking of answers in DeepQA"

- finale Stufe der pipeline zur Bestimmung der besten candidate answers

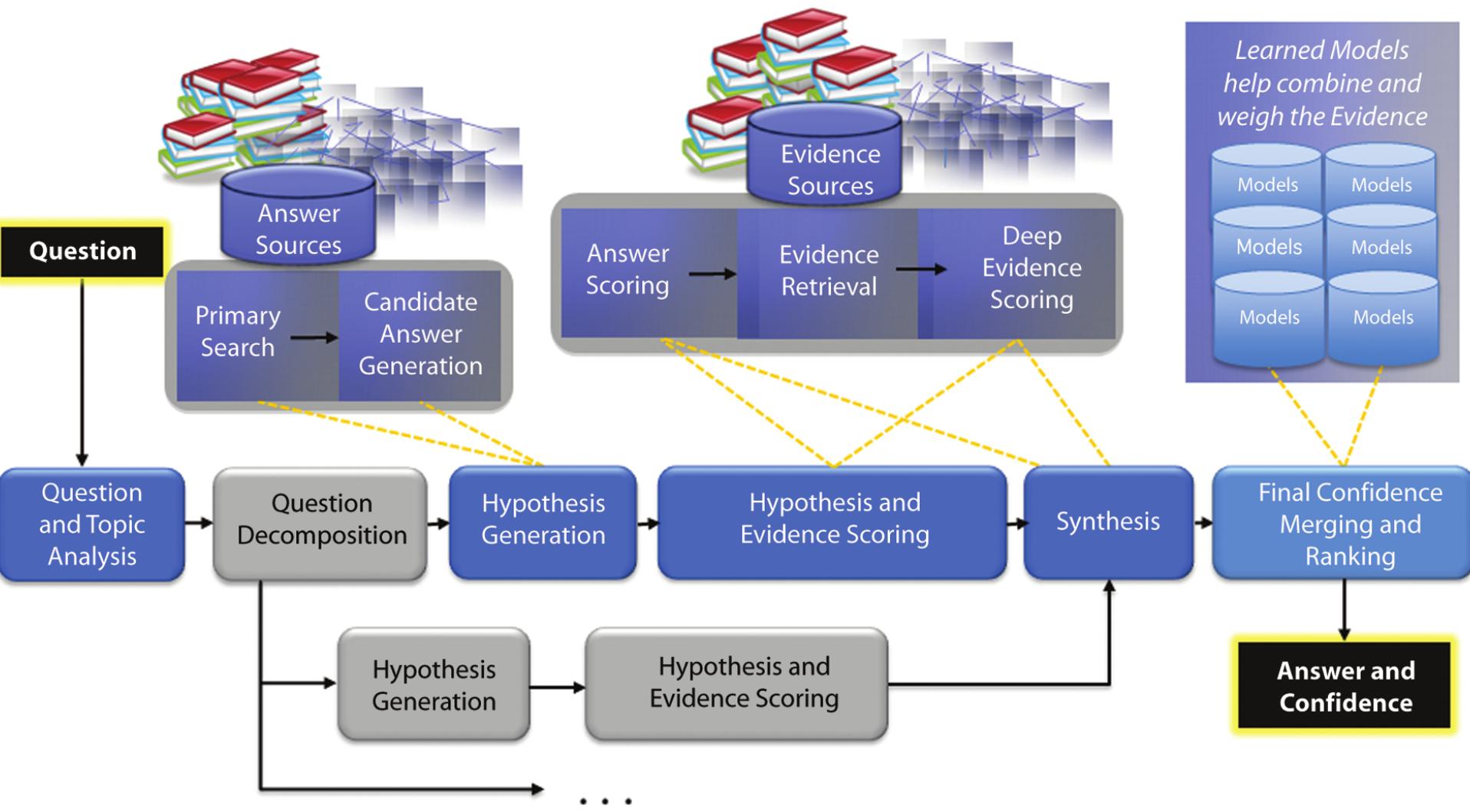
DeepQA Architektur



Massiv parallele Berechnung, Skalierung und Geschwindigkeit

Artikel 14 "Making Watson fast"

DeepQA Architektur



Massiv parallele Berechnung, Skalierung und Geschwindigkeit

- Latenz
 - Antwortzeit darf bei nicht mehr als 3 Sekunden liegen
- Einsatz paralleler Berechnung wo möglich
- Verteilung über asynchrone Nachrichten

Watson: Ein künstlicher Jeopardy! Spieler

- DeepQA selbst spielt noch kein Jeopardy
- zwei entscheidende Komponenten, um zum Jeopardy-Spieler zu werden:

Watson: Ein künstlicher Jeopardy! Spieler

Artikel 15 "Simulation learning and optimization techniques in Watsons game strategies"

- Handling der Spielstrategie

Watson: Ein künstlicher Jeopardy! Spieler

**Artikel 16 "In the game: The interface
between Watson and Jeopardy"**

- Antworten aussprechen
- Buzzer drücken
- Spielverlauf folgen

Zusammenfassung der Ergebnisse

- Drei Metriken
 - Präzision und "confidence"
 - Spielstärke
 - Performance der einzelnen Komponenten
- Watson answer-scoring baseline (WASB)
- Jeopardy-Herausforderung erwies sich als hervorragende Aufgabenstellung

Die nächsten Schritte - Watson 2.0

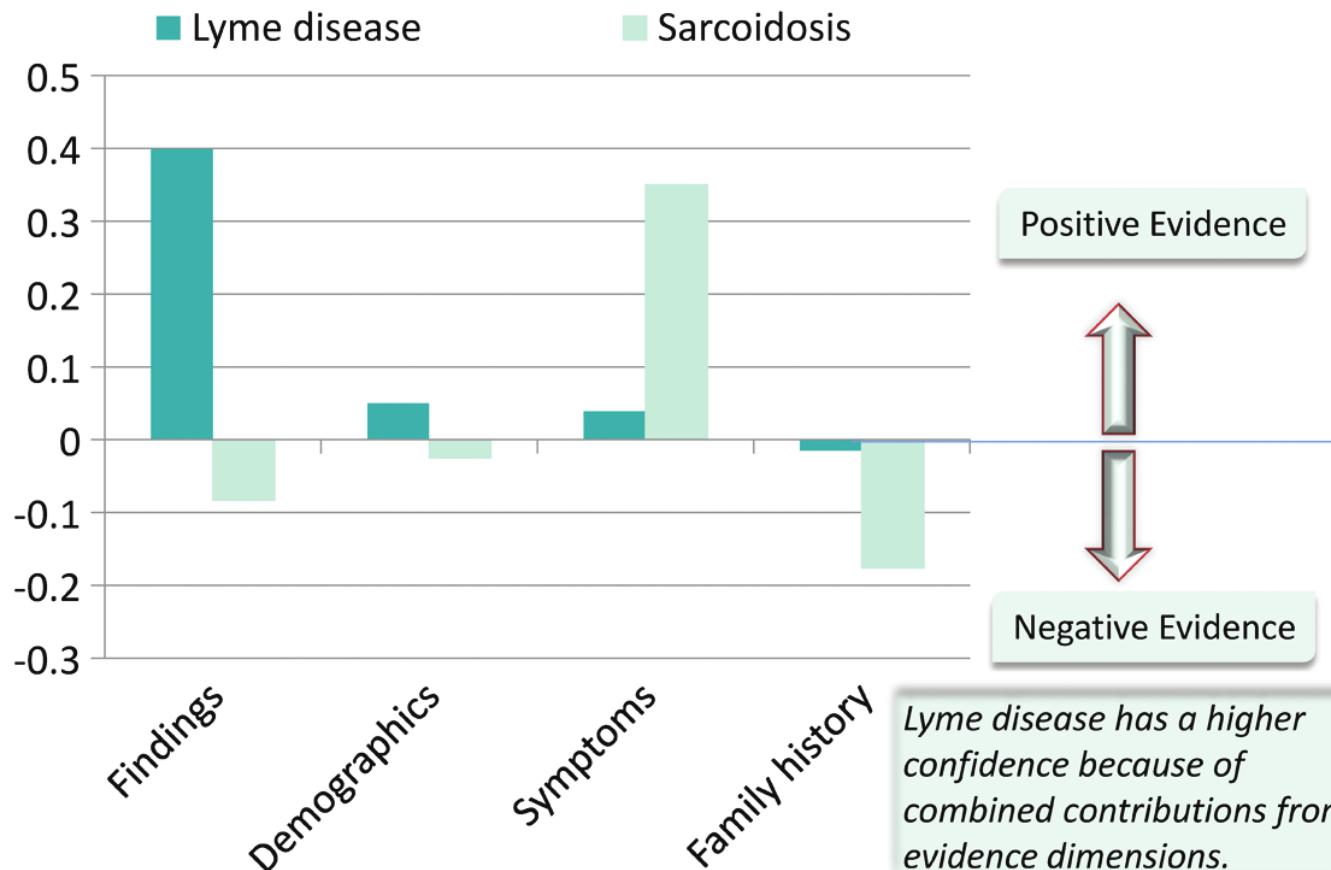
- Jeopardy repräsentiert ein eingeschränktes Interaktionsmodell
- Zukunft: in Dialog treten / "mixed-initiative-dialogue"
 - Nachfragen
 - Wissenslücken stopfen
 - Mehrdeutigkeiten auflösen
- Übergang von Beantwortung einzelner Fragen auf ganze Problemszenarien

Die nächsten Schritte - Watson 2.0

- sehr interessantes Einsatzgebiet:
- **Medizin**
 - die Patientenakte als Problemszenario
 - stetig wachsendes, unstrukturiertes Wissen
 - Fachbücher,
 - Journale,
 - Forschungsergebnissen, ...
 - Ärzte oft überfordert bei komplizierten Diagnosen

Differentialdiagnose durch Dr. Watson

Question: What are diseases, disorders, or causes of uveitis with circular rash, fever, headache, and family history of arthritis in a patient who lives in Connecticut.



"Merkblatt" :-)

- Pipeline-Architektur
- keine Komponente versteht die Frage komplett - auf das Zusammenspiel kommt es an
 - "Society Of Mind"

"Merkblatt"

- Trennung in:
 - Bestimmung von candidate answers
 - Hypothesenerzeugung
 - Suche von Beweisen für jede Hypothese
 - Gewichtung der Beweis-Scores
 - Ranking der aggregierten Scores pro Hypothese

"Merkblatt"

- Massiv parallele Berechnung
- Verarbeitung natürlicher menschlicher Sprache
 - stellt(e) große technische Herausforderung dar
- Sehr spannende Zukunftsaussichten