法律声明

- 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院和讲者共同拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

- 课程详情请咨询
  - 微信公众号：小象
  - 新浪微博：ChinaHadoop

第十讲: IBM Watson

Watson is ushering in a new era of computing

Watson


THINK
DENKE
$24,000
Who is Stoker?
((FOR ONE WELCOME OUR NEW COMPUTER OVERLORDS))
$1,000
$77,147
Who is Bram Stoker?
$ 17,973
$21,600
WHO IS BRAM STOKER?
$5600

Introduction to “This is Watson”

- History
- Deep QA pipeline and Watson components
- Results

What is open-domain QA?

- In information retrieval, an open domain question answering system aims at returning an answer in response to the user's question. The returned answer is in the form of short texts rather than a list of relevant documents.
- The system uses a combination of techniques from computational linguistics, information retrieval and knowledge representation for finding answers.

History of Jeopardy! & Watson

- Jeopardy! is a well-known television quiz show
- Research in open-domain QA requires advances in many areas of computer science and AI
  - information retrieval (IR)
  - natural-language processing (NLP)
  - knowledge representation and reasoning (KR&R)
  - machine learning
  - human-computer interfaces (HCIs)

Why use more unstructured data?

- With the enormous proliferation of electronic content on the web and within our enterprises, unstructured information (e.g., text, images, and speech) is growing far faster than structured information.
- Whether it is general reference material, textbooks, journals, technical manuals, biographies, or blogs, this content contains high-value knowledge essential for informed decision making.

Difficulties in using unstructured data

- Variable
  - there are many different ways to express the same meaning
- Polysemous
  - the same word or phrase may mean many things in different contexts
- Implicity
  - Unlike programming languages, human languages are not formal mathematical constructs.

Old framework

- UIMA
  - a software architecture and framework that provides a common platform for integrating diverse collections of text, speech, and image analytics independently of algorithmic approach, programming language, or underlying domain model.
  - provided the essential infrastructure needed to engage large-scale language understanding research.
  - facilitated the technical integration and scale-out of a broad range of analytic components, the challenge for building Watson became the development of the algorithms themselves and the architecture and methodology necessary for rapidly combining, measuring, and advancing them to succeed at Jeopardy!
- PIQUANT
  - A form of QA system. It was presumed a static predetermined set of answer types -- classes of concepts that the question was asking for.
  - Cons: there was no easy way to extend, rapidly measure, and advance with a diversity of techniques contributed by different researchers.

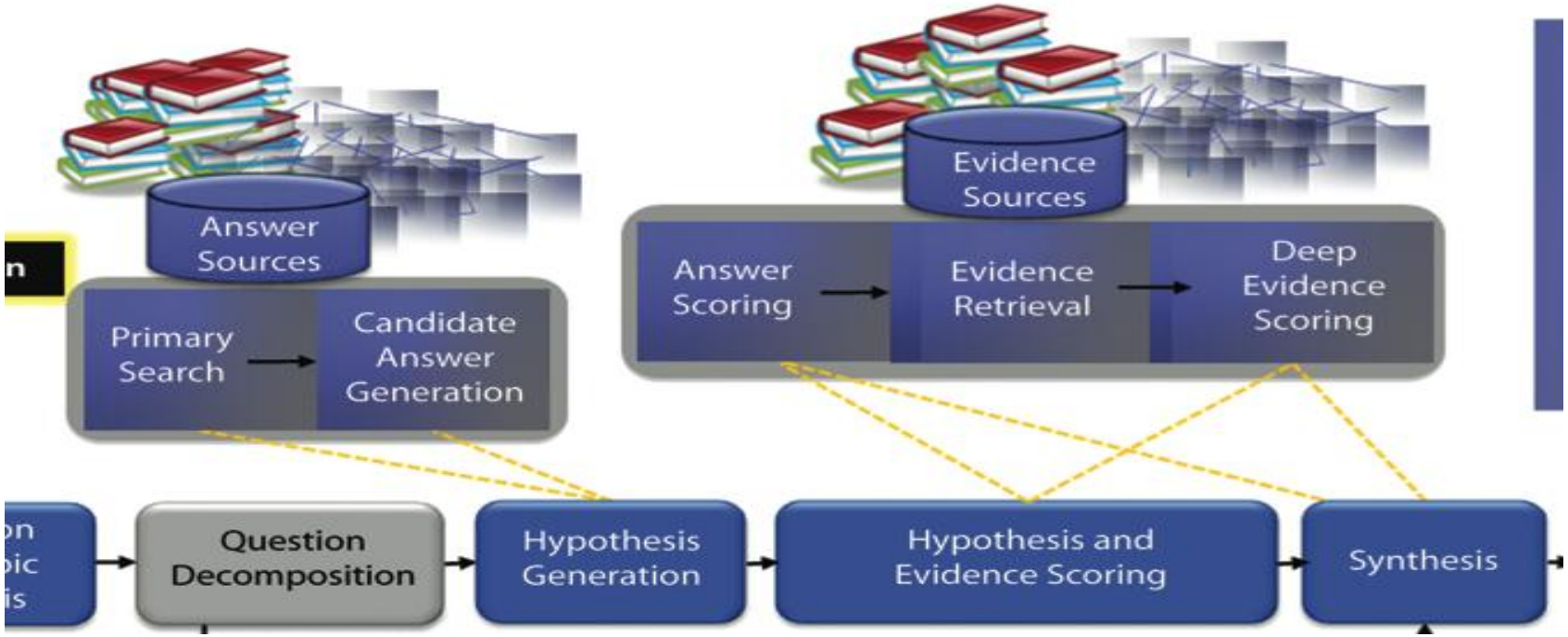
What it takes to win Jeopardy!

- Self-contained
  - No web search, no connection to the Internet, and no interaction with anyone else for help in understanding or answering the questions.
  - Analyze and store in its memory every bit of information it might consider during the game.
- Just three seconds to
  - Parse a clue
  - Understand what it was asking for
  - Relate its meaning to what it had Bread, determine the best answer, and compute whether it was confident enough to buzz in.
- Deeper analysis in
  - the clue
  - all the content it has read that might justify an answer
- 85% Precision@70
  - 85% precision at 70% answered to perform well in the game

Architecture of Watson

- Two important technical artifacts
  - Deep QA
    - An extensible software architecture for building QA systems
  - AdaptWatson
    - A methodology based on the ideas in [Towards the open advancement of question answering systems] for the rapid advancement and integration of many core algorithmic techniques
- The final result of the process - a ranked list
  - With candidate answers, each with a final confidence score representing the likelihood the answer is correct based on the analysis of all its supporting evidence.

DeepQA architecture



1. Understanding questions

- The more precisely Watson understands the clue, the better chance it has at finding and justifying answers.
- Get the lexical answer type

RECENT HISTORY: President under whom the U.S. gave full recognition to Communist China. (Answer: "Jimmy Carter")

FOLLOW THE LEADER TO THE CAPITAL: Pervez Musharraf. (Answer: "Islamabad")

How to analyze the category of this ?

2. Answer and evidence sources

- Started with the obvious - encyclopedias and reference books - and then developed a semiautomatic process for growing Watson's content.
- Several tradeoffs to deal with
  - Adding content -> may contributes noise rather than value -> hurt performance of the system.
  - Hardware is limited to add much content.
- Some questions
  - How raw textual content was selected, evaluated, and automatically expanded
  - The approach to analyzing the raw content for building a derived resource

3. Discovering candidate answers

- DeepQA does not assume that it ever completely understands the question with respect to any preexisting model, set of type ontologies, or database schema.
- It parses results and question analysis result in multiple interpretations (different queries).
- These queries are run against different structured and unstructured sources .

3. Discovering candidate answers

- first generate a broad set of candidate answers
  - Each candidate answer combined with the question represents an independent hypothesis.
  - Each becomes the root of an independent process that attempts to discover and evaluate supporting evidence in its candidate answer.
- Metric: candidate binary recall (precision & recall)
  - the percentage of questions for which the correct answer is generated as a candidate
  - maximizing candidate recall at the hypothesis generation phase of the DeepQA pipeline.
- The rest part of pipeline is finding the best answer and computing an accurate probability that it might be correct.

4. Right type

- An important class of evidence considered is whether the answer is of the right answer type.
- Difficulties
  - PIQUANT anticipated all answer types and built specialized algorithms for finding instances of anticipated types – would not be sufficient.
  - Jeopardy! has too many types that a priori information extraction components for each one is impossible.
  - Those types' meanings are highly contextual and difficult to anticipate.

4. Right type (cont)

- Type coercion
  - much more dynamic and flexible
  - heavily relying on the context in the question
  - It takes a lexical answer type such as “direction” and poses a type hypothesis for each candidate.
  - example
    - is_a (down, “direction”)
    - is_a (grain, “direction”)
  - then consults a wide variety of structured and unstructured resources using a diverse set of algorithmic techniques to gather evidence for and against the type hypothesis.
- No component in the system was pretrained in classifying
- Watson uses an extensible ensemble of techniques based on a wide variety of textual and ontological resources
  - PRISMATIC
  - YAGO
  - WordNet

5. Collecting and scoring evidence

- collect and score additional evidence to support the answer generated before.
- Design algorithms
  - to produce a confidence score that indicates the degree to which a piece of evidence supports or refutes the correctness of a candidate answer
- Grammar-based techniques
  - address syntactic and shallow semantic features of language
- Relation extraction techniques
  - look deeper into the intended meaning, attempting to find semantic relationships

broad-domain relation extraction and scoring

- Two methods
  - manual pattern specification (rule based)
    - Pros: more precise and used by many components
    - Cons: need manual effort to develop patterns
  - statistical methods for pattern elicitation
    - Pros: automatically learn how to extract semantic relations from the training data.

MOTHERS & SONS: Though only separated by one year in real life, she played mother to son Colin Farrell in "Alexander." (Answer: "Angelina Jolie")

- Relation: starring (she, "Alexander")

Example

LANGUAGE: The lead singer of the band Dengue Fever is from this country & often sings in Khmer.
(Answer: “Cambodia”)

Assume the country being asked for is a country whose national language is Khmer

PRESIDENTS: The only 2 consecutive U.S. presidents with the same first name.
(Answer: “James Madison and James Monroe”)

Enumerates all entities of the given type (U.S. presidents and countries) and looks up the relevant characteristics.

6. Special questions

- There are small part of questions in Jeopardy! with special attributes.
- Handled by special algorithms.
  - first detect their occurrence
  - then to favor the algorithmic paths in the DeepQA architecture specialized for finding and synthesizing answers.

Other puzzles

- Implicit relationships extraction
  - Build the ability to find what different concepts have in common
- Breaking the question down
  - Breaking a question down into logical subparts, so that the subparts may be independently explored and the results combined to produce the answer.

Fact-Based Question Decomposition in DeepQA, Kalyanpur et al.

Identifying Implicit Relationships, Chu-Carroll et al

7. Merging evidence and combining confidence

- Watson may consider 100 candidate answers for some question.
- Each evidence-answer pair may be scored by 100 independent scorers.
- Each scoring algorithm produces a confidence.

- Final stage in DeepQA: (figure out the best way to) combine the confidence of many different scorers across different pieces of evidence to produce a single probability for each candidate answer.
- using a statistical machine learning framework.

Merge the score

- learning approach
  - machine learning is used to weigh and combine scores
  - evidence scoring algorithms can be easily introduced, revised, and reconfigured without the need for an error-prone manual effort to determine how to combine the various evidence scores.

A Framework for Merging and Ranking of Answers in DeepQA, Gondek et al.

Massive parallelism, scale-out, and speed

- Latency
  - the time it took to answer a question
- scaling DeepQA by embedding it in UIMA-AS
  - UIMA-AS allows any UIMA application to be deployed as a collection of asynchronous processes that use standard messaging infrastructure to communicate.
- implement an embarrassingly parallel process from the beginning

Making Watson Fast, Epstein et al.

Play the game

- Two principal components
  - to handle the game strategy
  - to interface with the game itself
- Watson could not see or hear.
  - The team focused on the natural-language QA task and opted not to address visual or auditory clues.

In the Game: The Interface between Watson and Jeopardy!, Lewis

Result summary

- Metrics in measuring the results
  - precision and confidence
  - game-winning performance
  - component performance.

Metric 1

- precision and confidence
  - represented by plotting QA precision against percentage answered to produce a confidence curve.
  - running DeepQA on 200 games' worth of blind data which contain 12,000 questions that were never viewed by the developers and never used to train the system.
  - Compared with the records of two Jeopardy! champion players.

Metric 2

- game-winning performance
  - includes the impact of strategy and speed in formal real-time gameplay with original Jeopardy! equipment
  - Watson played 55 real-time previously unseen games against these players and won 71% of them.

Metric 3

- The quantity and diversity of components used to implement DeepQA make it extraordinarily robust and flexible
- The system does not heavily depend on any one of them.
- Watson answer-scoring baseline (WASB) system
  - includes only one evidence-scoring component
  - evaluated the impact of adding various components to the WASB system
  - Why so? Because ablation from the whole system does not provide the meaningful insights for us.

Future directions

- Watson 2.0
  - strikes a balance somewhere between a search system and a formal knowledge-based system.
  - consider not simply queries but also entire problem scenarios
  - produce hypothetical solutions and compelling evidence
  - automatically consuming and analyzing natural-language sources of knowledge
  - can produce evidence profiles and supporting evidence
  - will support mixed-initiative dialogue


Watson is creating a new
partnership between people
and computers that **enhances**,
scales and **accelerates** human
expertise.

Medical information is growing faster than clinicians are able to put it to use

Complexity

MDs must consider 100 variables per case by 2020 and 1000 by 2025

Evidence

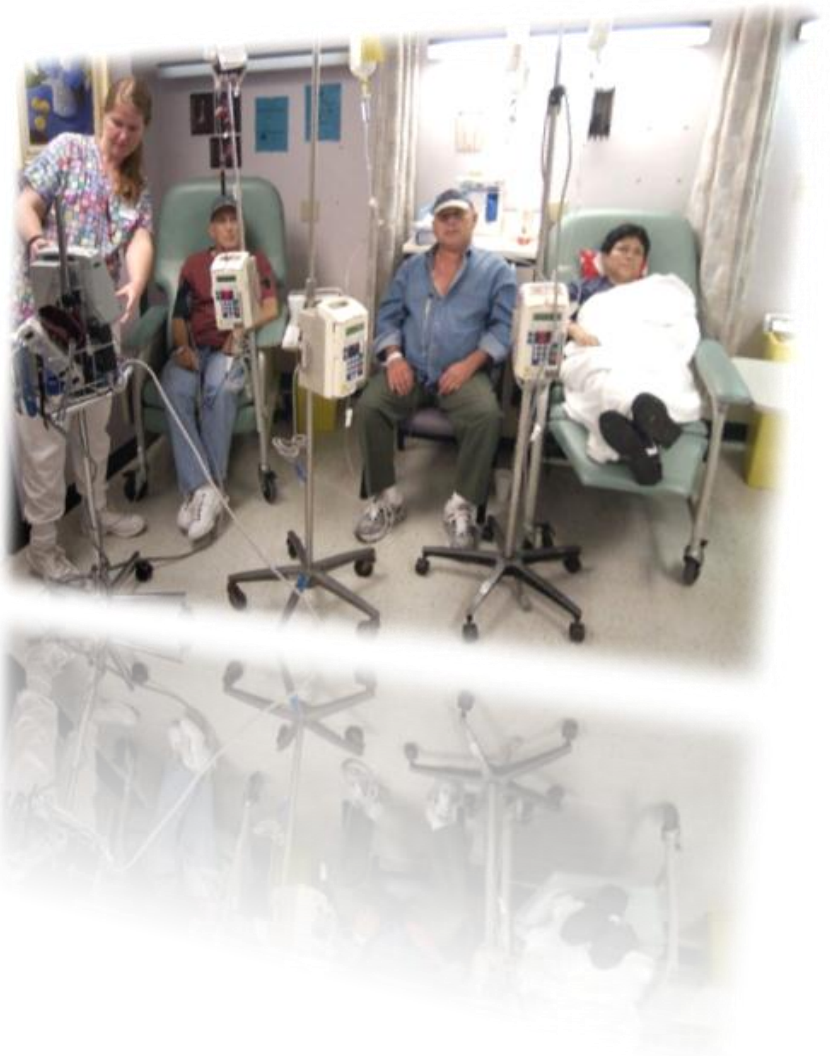
<50% of medical decisions are evidence-based

Data

80% of medical info is unstructured making it difficult for systems to use

Time

Demand for services to grow 42% by 2025 → ~1500 oncologist shortage



Watson Discovery Advisor represents a new approach for Oncologists

Evidence-based test and treatment suggestions

Drawn from 600K+ pieces of evidence and 2M pages of text from 42 publications

Expert training by Memorial Sloan Kettering Oncologists

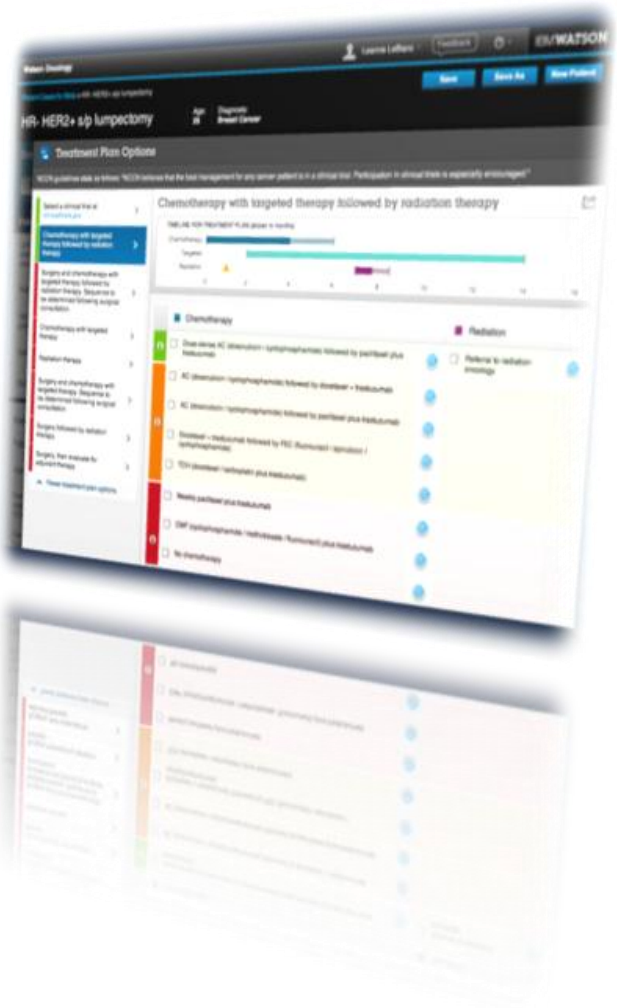
5000+ physician and analyst training hours

Evolves with the fast-changing field

Improves over time as a learning system with usage and training

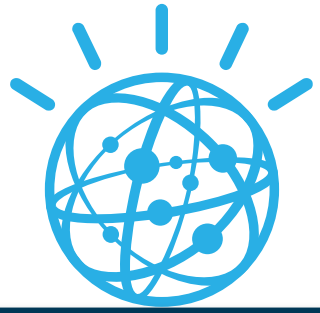
Full transparency into sources behind suggestions

Including journal articles, physicians' notes, NCCN guidelines and best practices



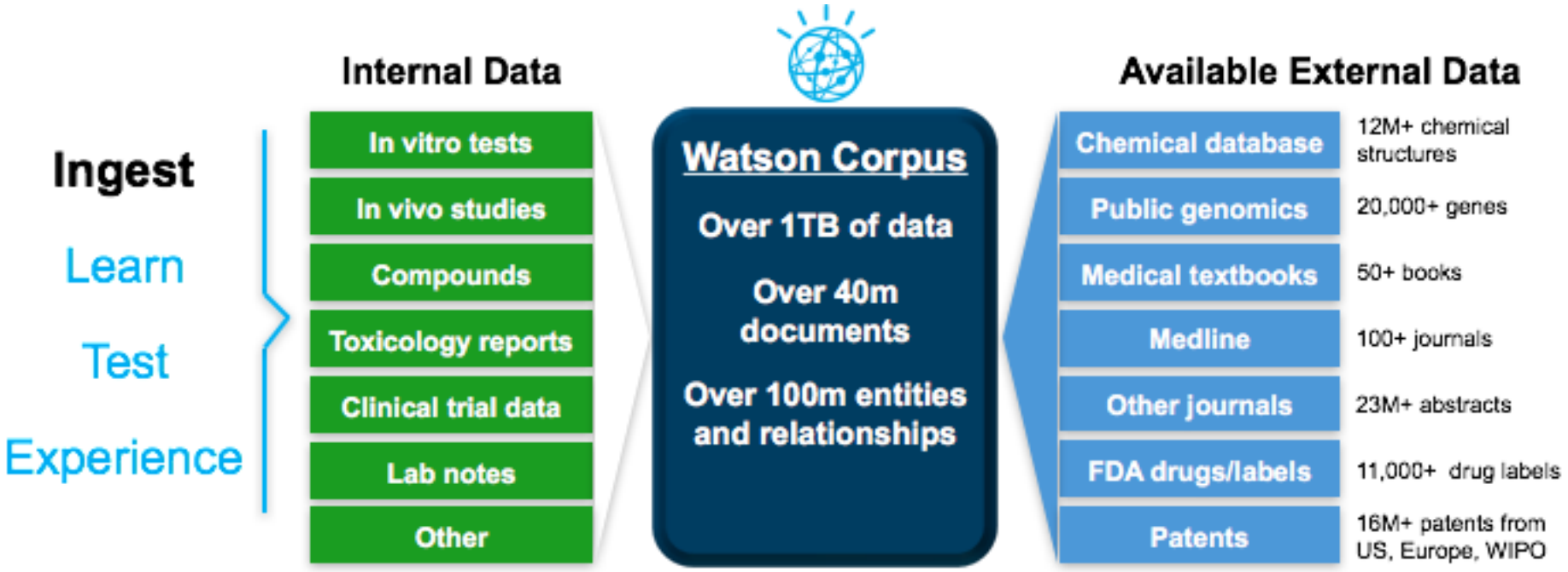
Watson follows a similar development path as expertises

Study
Apprentice
Practice
Master

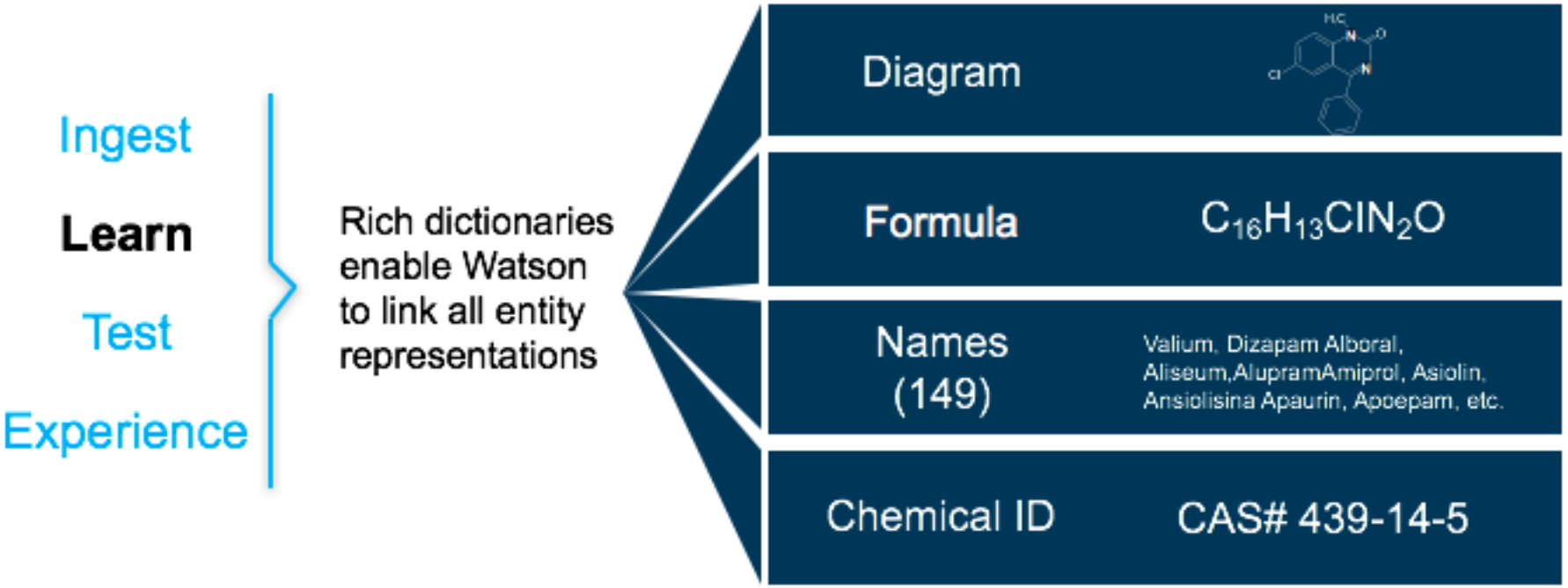


Ingest
Learn
Test
Experience

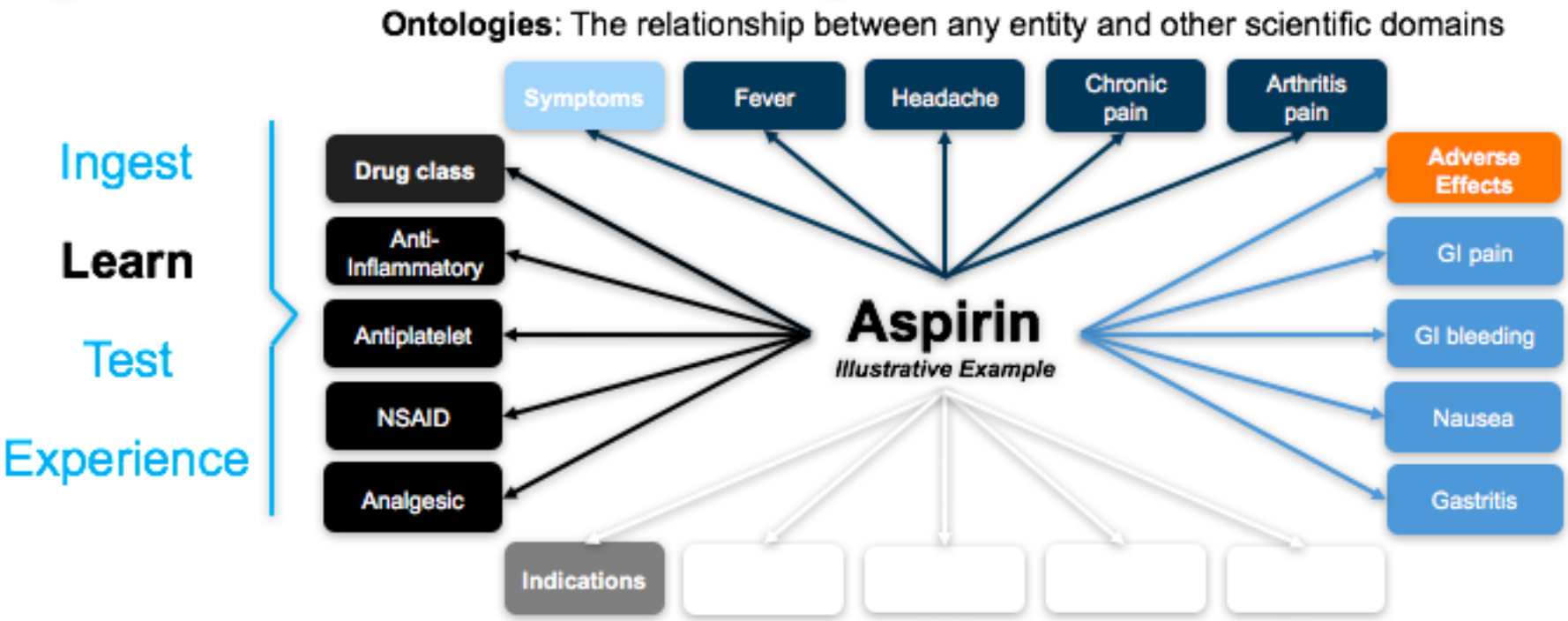
Watson enables insights by connecting and analyzing hundreds of internal and external data sources in minutes rather than weeks



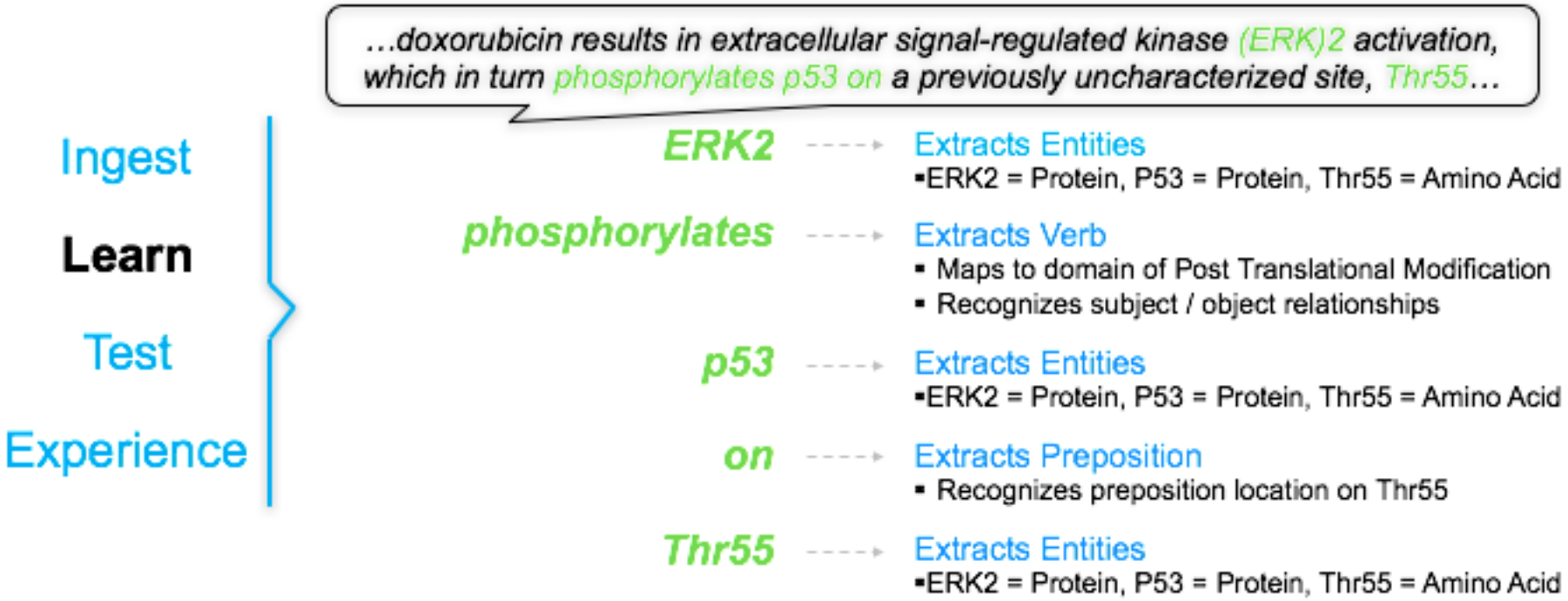
Not just a search engine, Watson understands and interprets the language of science



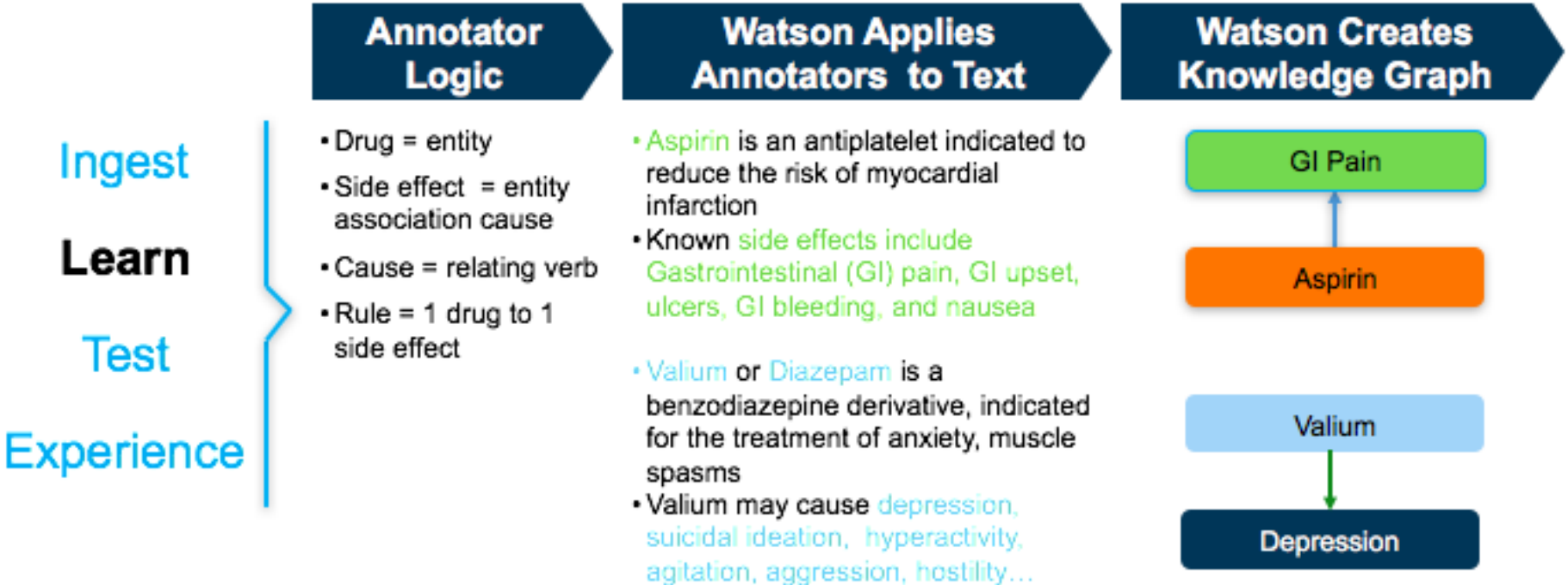
More than mere text mining, Watson can identify relationships



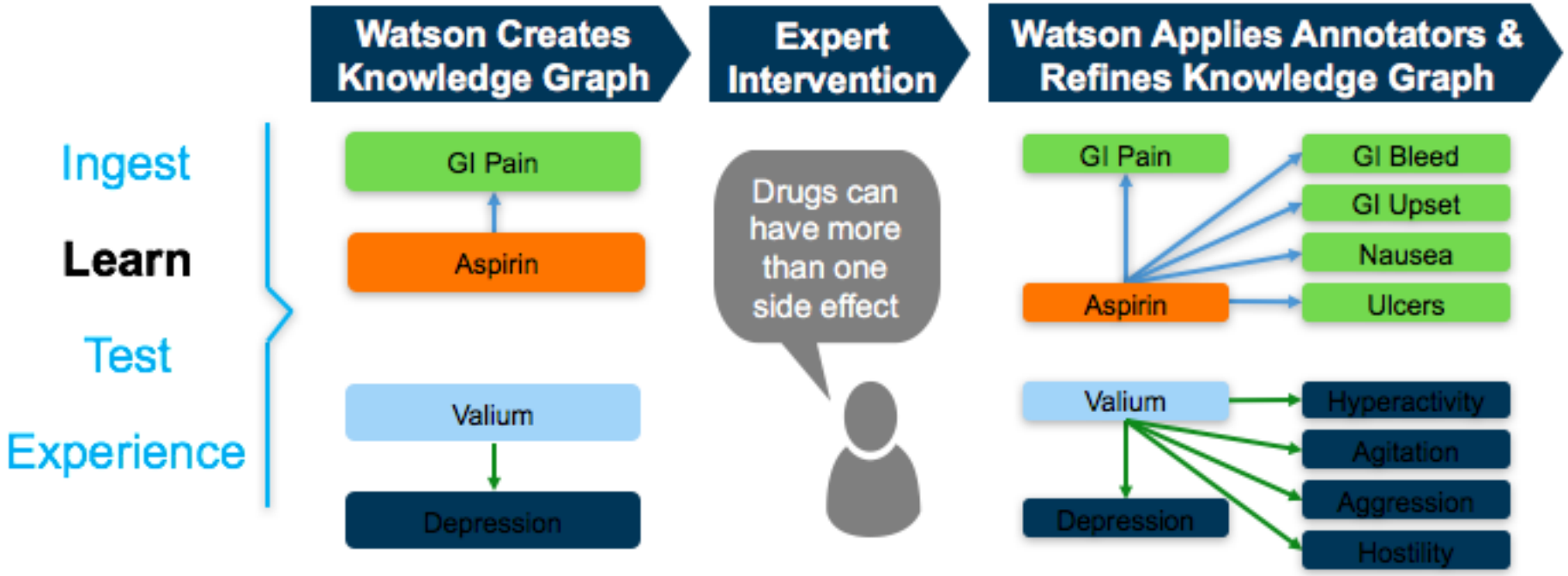
Annotators allow Watson to read and extract appropriate information



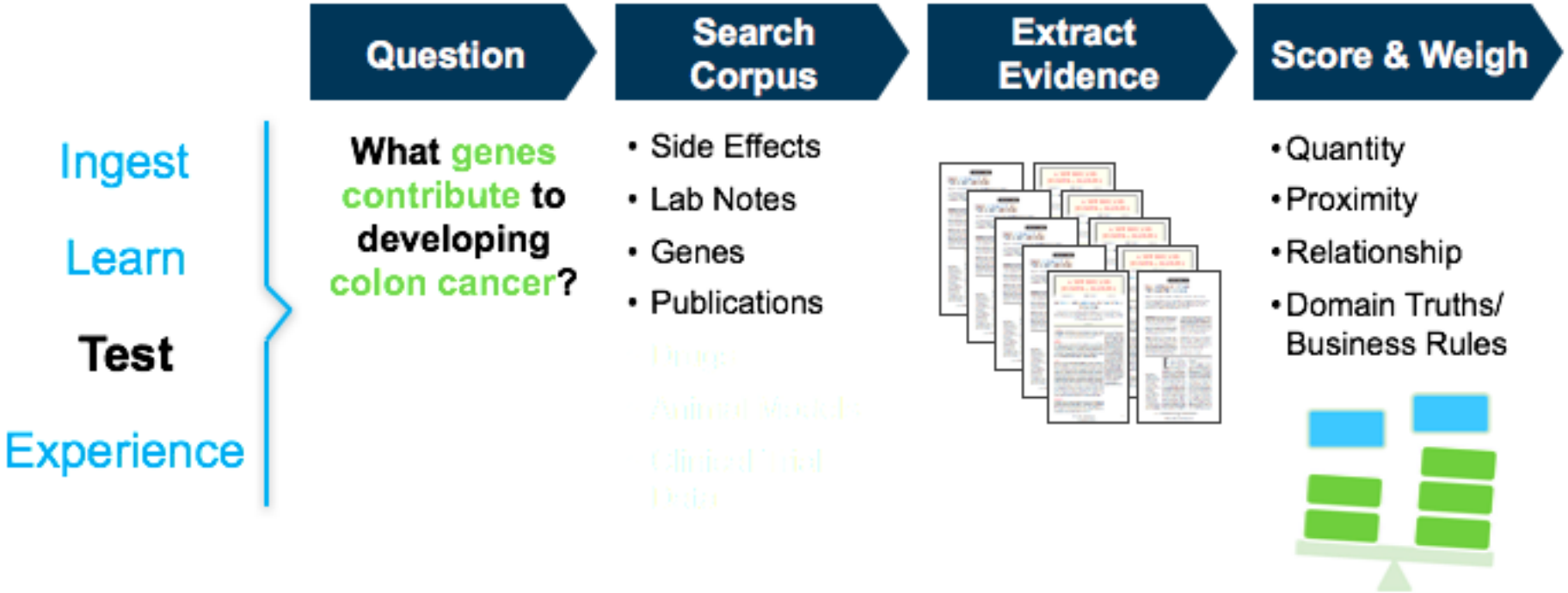
Creating the Knowledge Graph



Expand the Knowledge Graph by Learning from Experts



Beyond mere algorithms, Watson evaluates supporting evidence



The Result: Watson enables breakthrough insights after analyzing thousands of articles and other corpus data in minutes

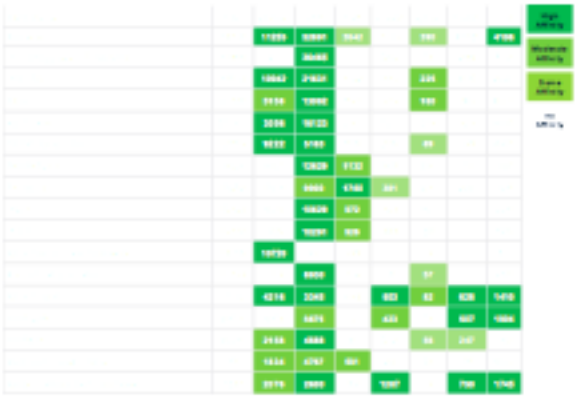
Ingest

Learn

Test

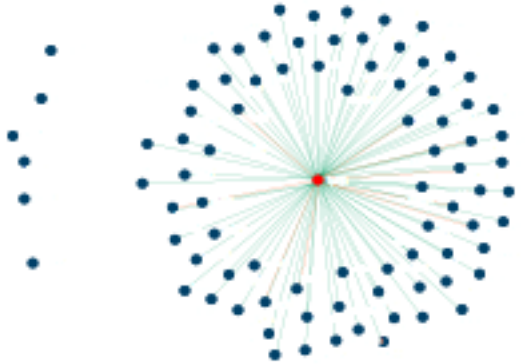
Experience

Co-occurrence Table



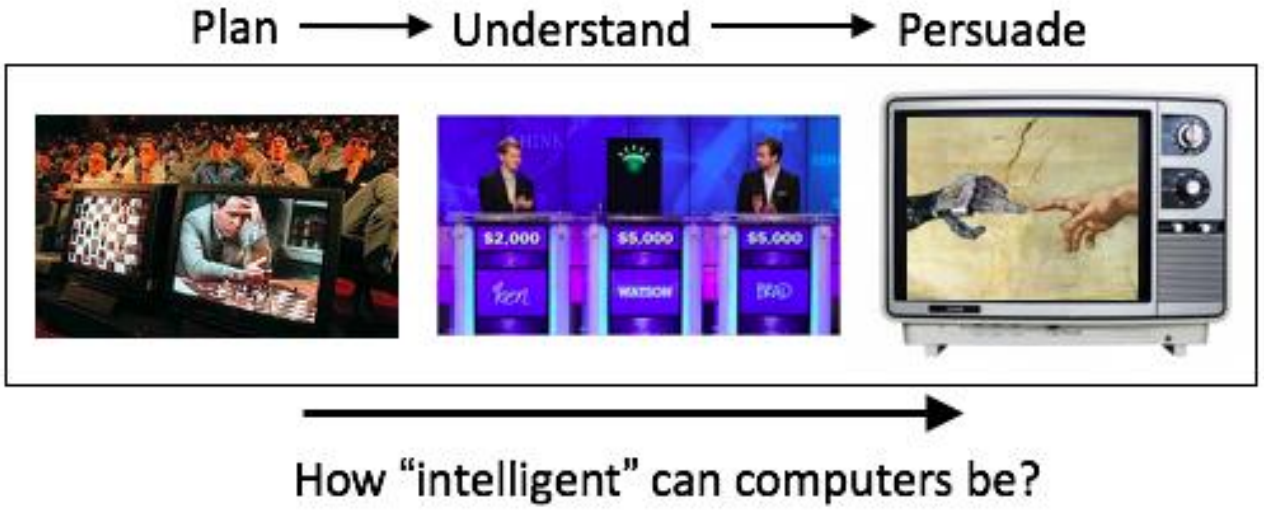
- Select entities from two different ontologies (i.e. disease/gene)
- Visualize co-occurrence
- Use statistics to spot the intersections
- Drill down to see the evidence

Gene Network

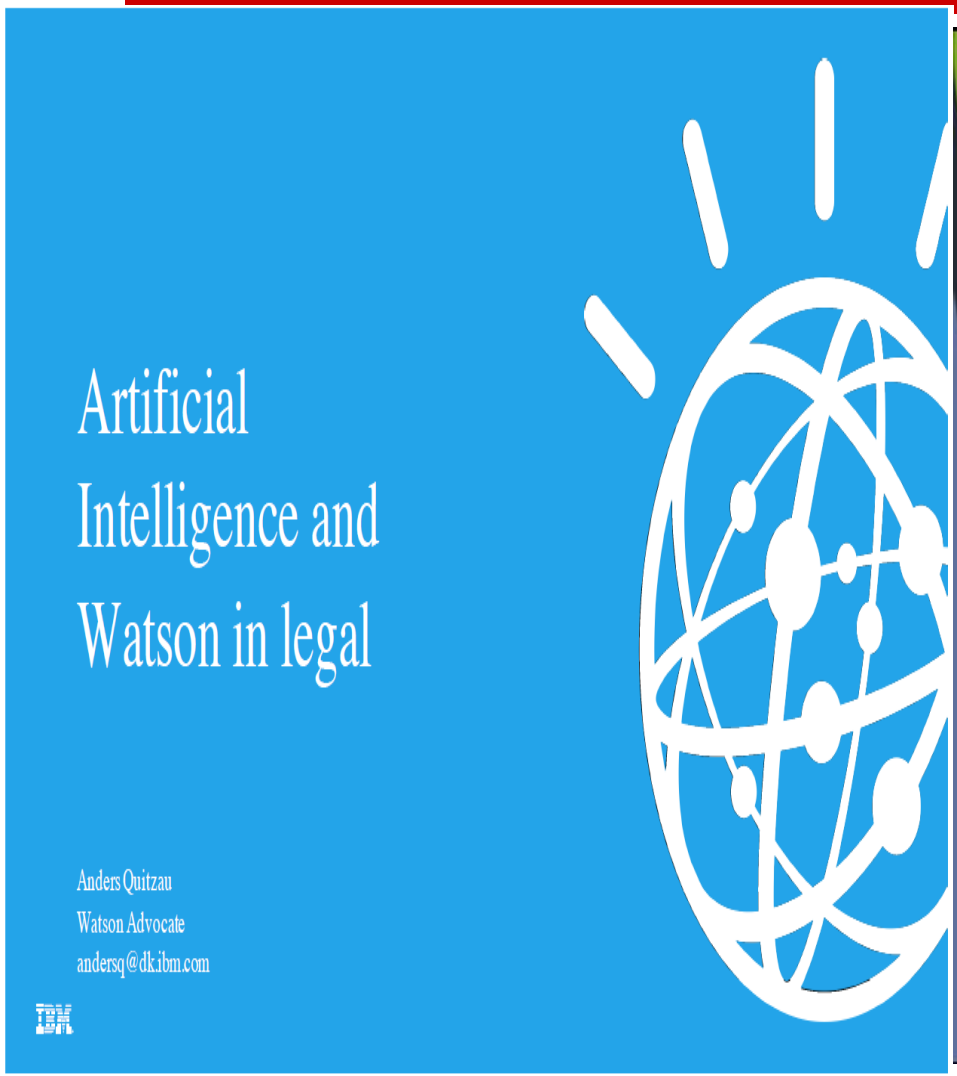


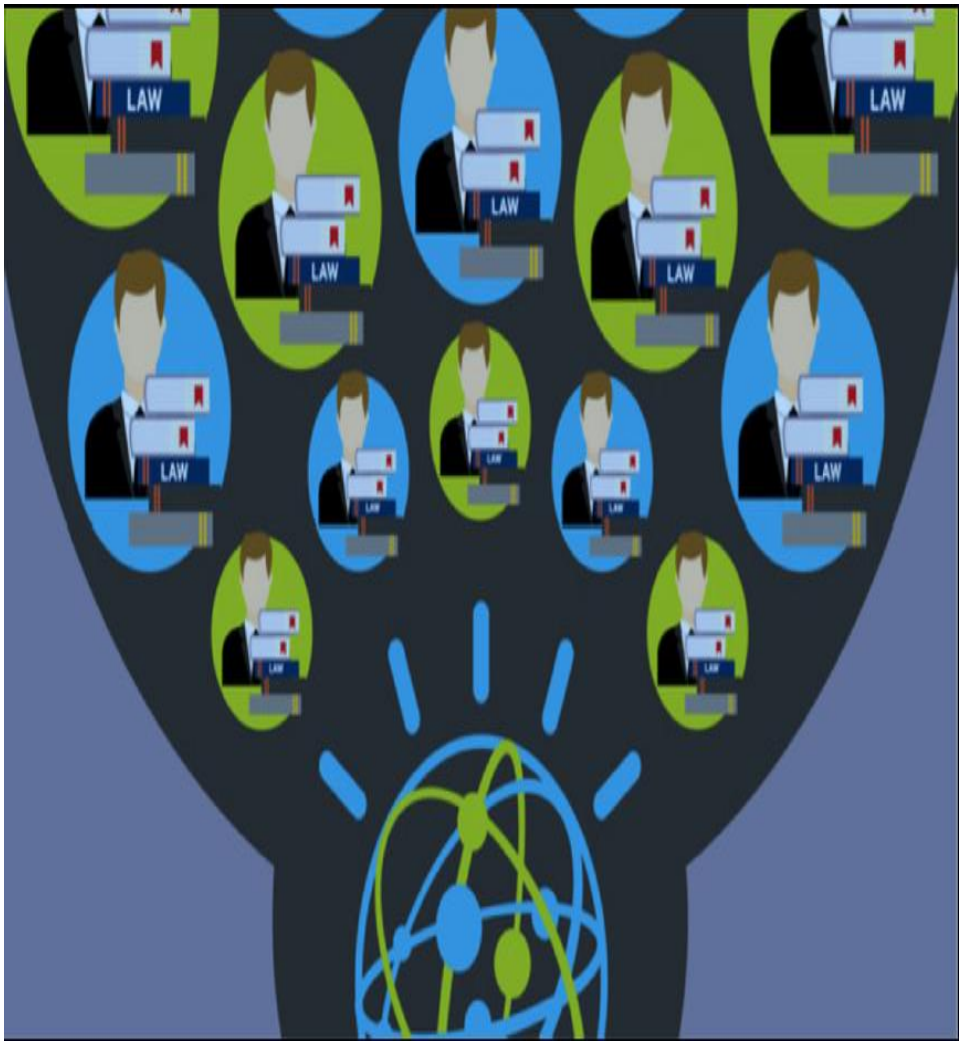
- Select two or more genes of interest
- See network of relationships
- Show strength, nature & proximity of the relationship
- Colored vectors indicate the nature of the interaction
- Hover over connections to see the evidence

What's Next?


Plan
Understand
Persuade
$2,000
$5,000
$5,000
How “intelligent” can computers be?

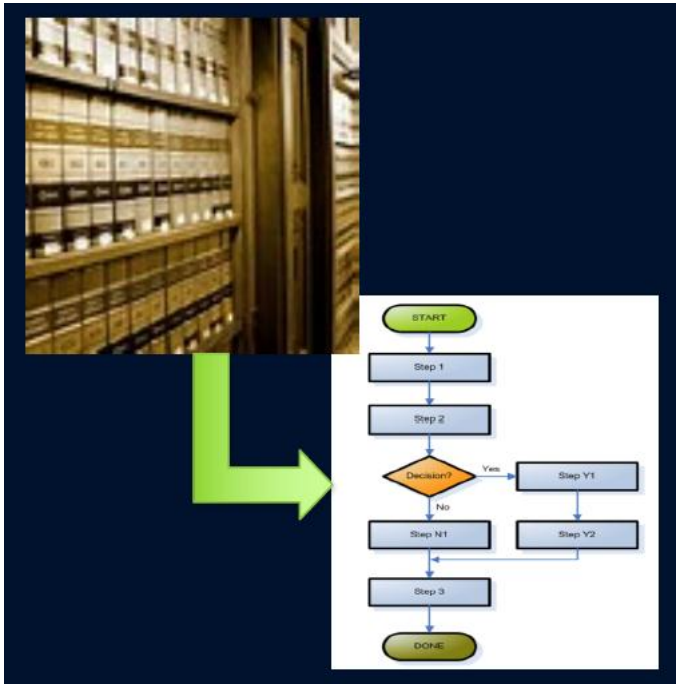
Watson role in the legal industry





Analytics are Necessary to Optimize Operations

- **Data Driven:**
- Most industries have adopted the use of data to drive optimized business process
- **Hidden Data:**
- The legal industry is unique in its scarcity of quantitative data that can be used to drive data optimized decisions
- **Cost:**
- Current methods of getting the data rely primarily on manual extraction or the creation of large, brittle rule sets
- **Opportunity:**
- Automate the extraction of quantitative data for use in optimized data driven decision making



Train Watson to build structured data sets

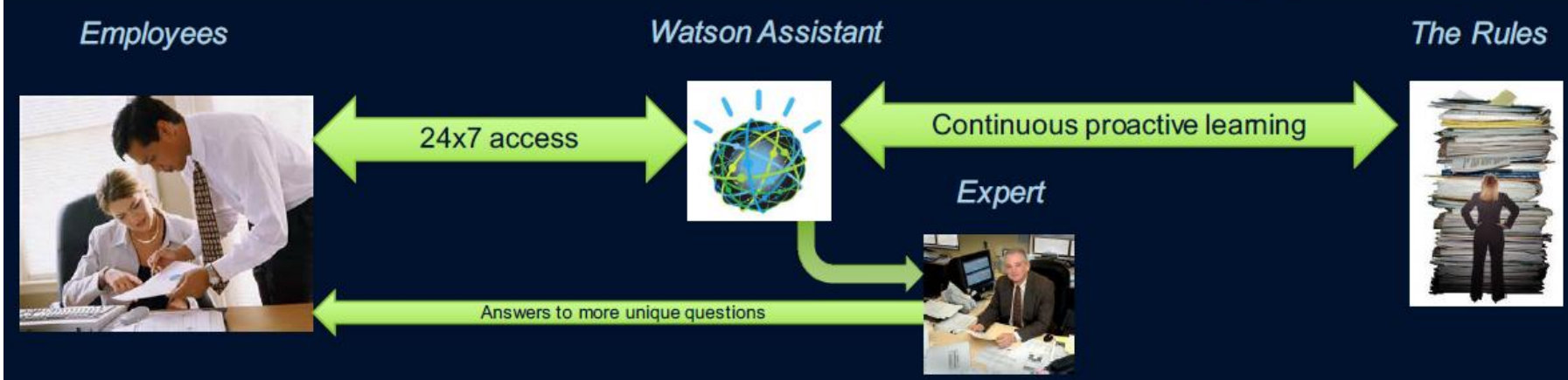
We can train Watson to review all incoming documents and search for the key data features we need to do discovery and prediction.

With structured data sets we can determine things like:

- Likely cost of matter
- Likely value of matter
- Likely outcome and key predictors
- Hidden correlations between data features

Watson will transform compliance environment

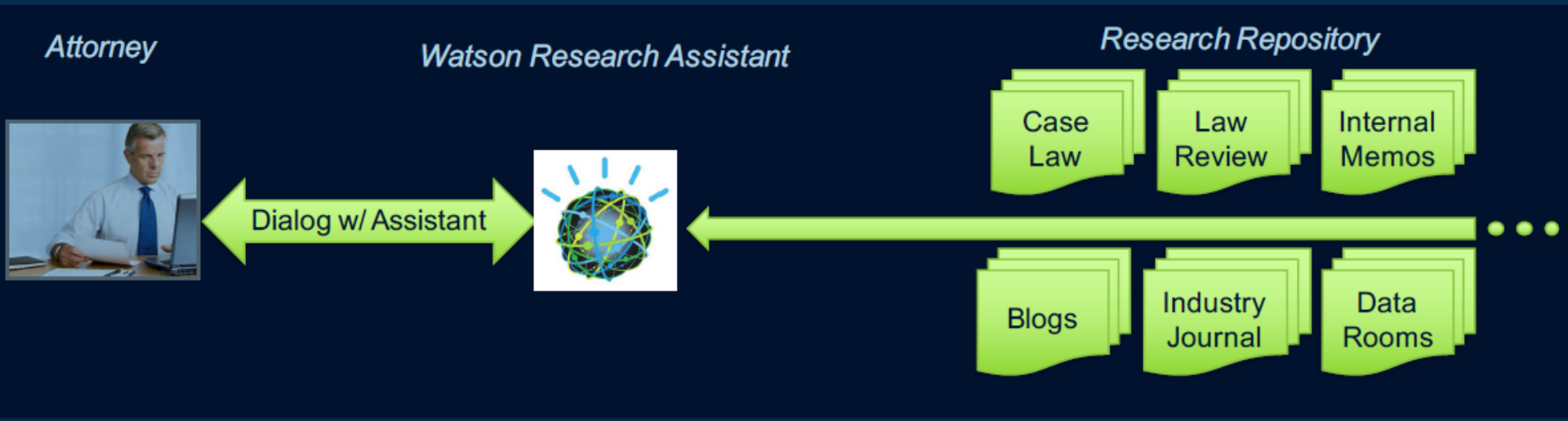
- **Accurate:** *Watson is trained on the actual content providing authoritative answers*
- **Evolving:** *Watson keeps up with changes to rules and interpretations*
- **Accessible:** *Watson is deployed directly to the point of interaction through the "right" technology at the right time*



- **Efficient:** *Watson deflects many questions, allowing people to act on the "interesting" ones*
- By getting accurate information in the hands of the people making important decisions at the time they are making decisions, we reduce the cost of both compliance and non-compliance

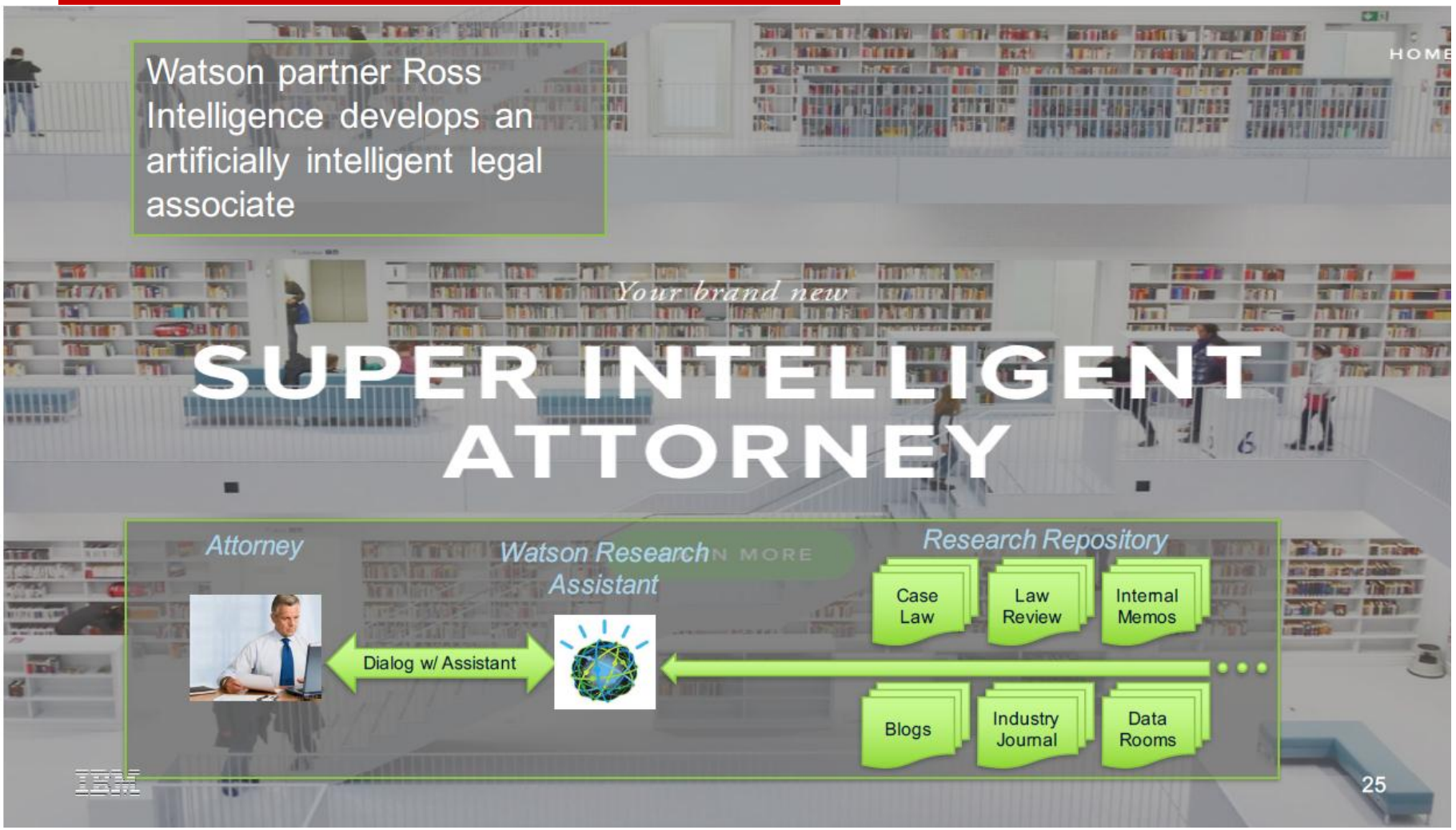
Watson can perform comprehensive legal research – faster

- **Accessible:** *Watson understands and interacts in natural language*
- **Comprehensive:** *Watson includes all sources simultaneously*
- **Fast:** *Watson is designed to work in real time on massive amounts of data*

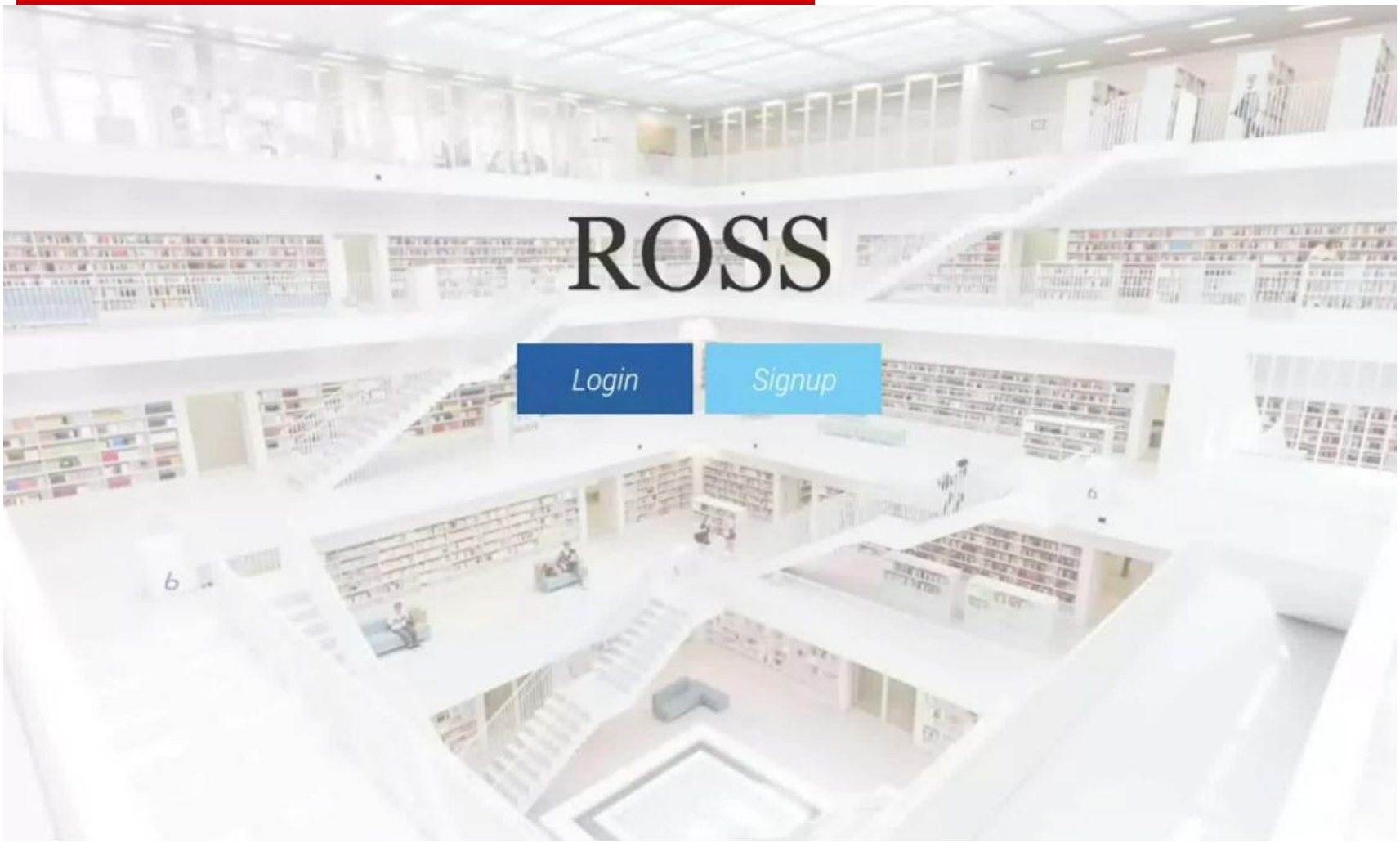


- Watson transforms legal research from an extended, manual and key word based process to a real-time, free flowing, natural language based dialog that enables the attorney to interactively explore and ultimately decide on a course of action

Watson


Watson partner Ross Intelligence develops an artificially intelligent legal associate
Your brand new
SUPER INTELLIGENT ATTORNEY
Attorney
Watson Research Assistant
Research Repository
Case Law
Law Review
Internal Memos
Dialog w/ Assistant
Blogs
Industry Journal
Data Rooms
25

Watson



Watson

"If a business is not reinventing itself to changing market conditions then it is highly likely it will go into decline or be taken over by those that are better adapted to the new environment. This statement is no less true for law firms than for any business"

The Law Society of England and Wales – Future's Panel

IBM

谢谢大家！

本课程课件由OpenKG提供支持

联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：大数据分析挖掘
- 新浪微博：ChinaHadoop