Open Research Questions:
What Do Knowledge Graphs Really Know?

Mark A. Musen, M.D., Ph.D.

Stanford University
musen@Stanford.EDU
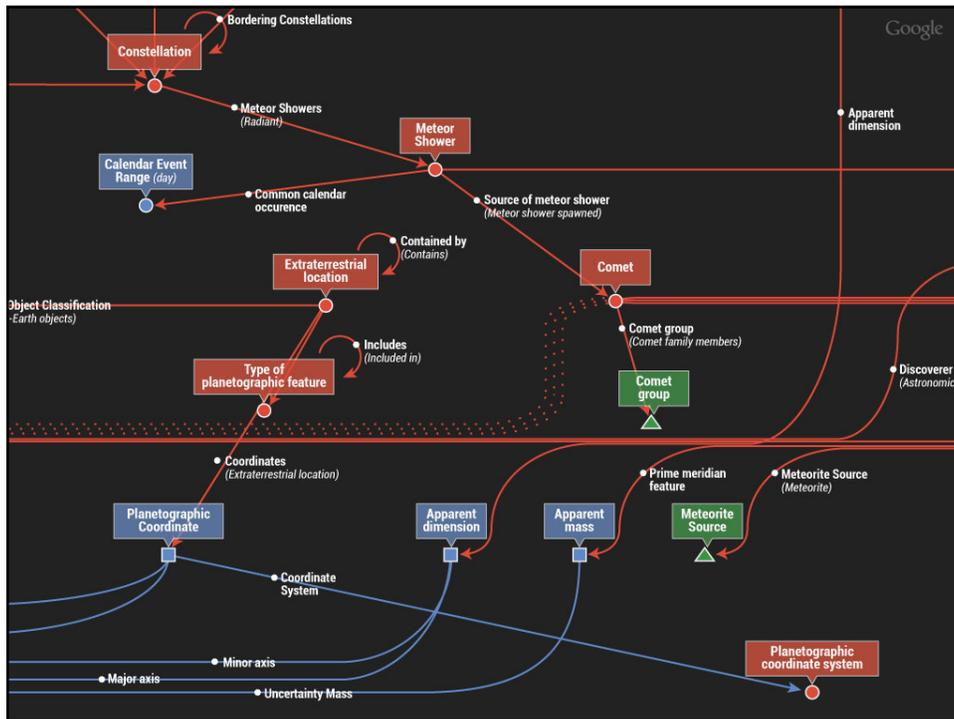

BMIR
Stanford Center for
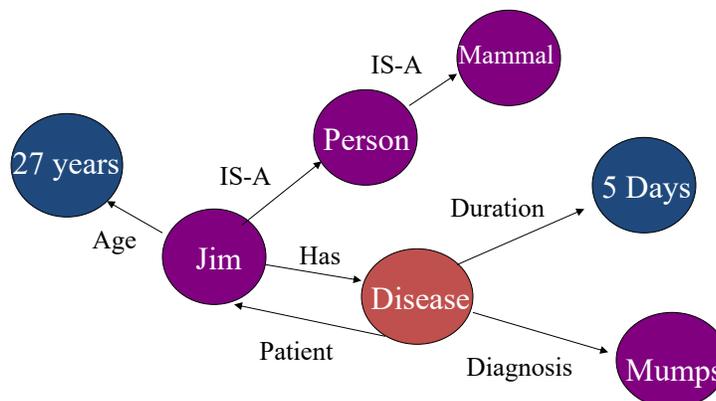Biomedical Informatics Research

CONNECTING DATA TO HEALTH

1

2

What have we been talking about in this seminar?

- We all understand what a graph is
- But what do we mean by *knowledge*?
- And what does it mean to say that we have put knowledge in a graph?

3

A Semantic Network (ca. 1970)

A directed graph where vertices are concepts and edges are relations

4

Then the expert-system craze started at Stanford …

- Stanford wanted to help doctors make better decisions regarding the empirical treatment of patients with severe infections
- Researchers considered the use of semantic networks, but they wanted more advanced reasoning
- The result was MYCIN, one of many **rule-based systems** to emerge in the 1970s and 1980s

5

Sample MYCIN Rule

RULE 579

If: 1) The infection that requires therapy is meningitis,
           2) The patient's chest x-ray is abnormal, and
           3) Active-Tb is one of the diseases that the
               patient's chest x-ray suggests

Then:
   There is strongly suggestive evidence (.8) that
   Mycobacterium-Tb is one of the organisms
   (other than those seen on cultures or smears)
   that might be causing the infection

Author: Yu

6

Sample MYCIN Case

30) Is Case 168 a compromised host (e.g., alcoholic, sickle-cell-disease, immunosuppressed)?

** NO

31) For how many days has Case 168 had abnormal neurologic signs?

** 1

32) For how many days has Case 168 had neurological symptoms?

** 9

Has Case 168 had recent evidence of tender or enlarged salivary glands?

** WHY

7

Sample MYCIN Case

My preferred therapy recommendation is as follows:

Give the following in combination:

1) ETHAMBUTAL

Dose: 1.28g (13 100mg tablets) q24h PO for 60 days [calculated on basis of 25 mg/kg] then 770 mg (7.5 100mg tablets) q24h PO [calculated on basis of 15 mg/kg]

Comments: Periodic vision screening test are recommended for optic neuritis

2) INH

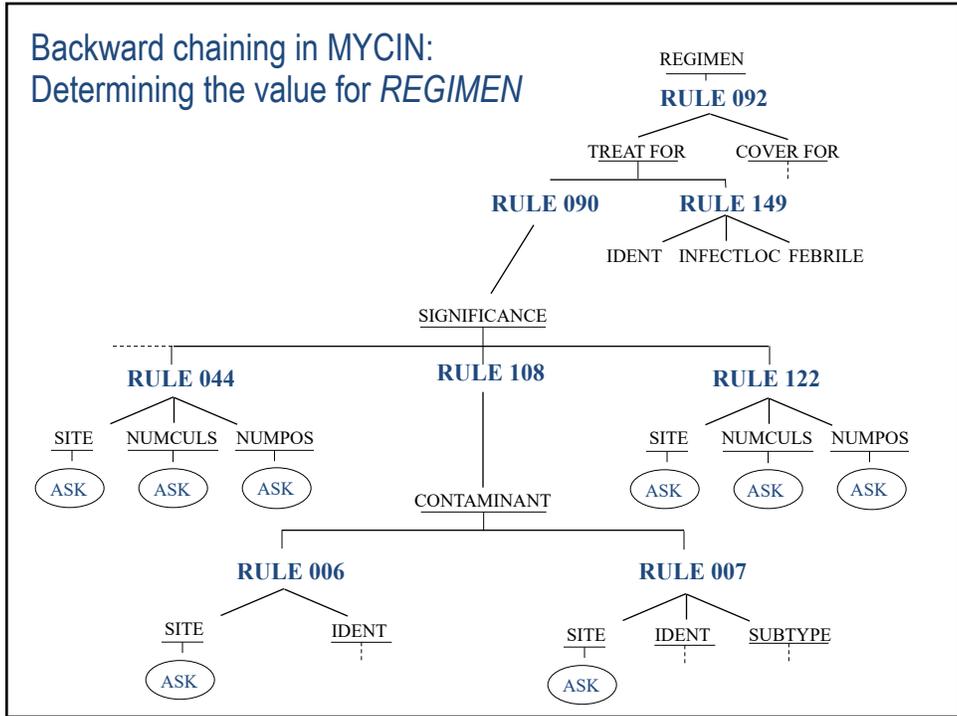Dose: 513mg (5 100mg tablets) q24h PO [calculated on basis of 10 mg/kg]

3) RIFAMPIN

Dose: 600 mg q24h PO

Comments: Administer dose on an empty stomach

8

Backward chaining in MYCIN: Determining the value for *REGIMEN*

```
                                    REGIMEN
                                    RULE 092
                         TREAT FOR          COVER FOR
                    RULE 090                RULE 149
                         |             IDENT  INFECTLOC  FEBRILE
                    SIGNIFICANCE
         RULE 044        RULE 108            RULE 122
    SITE  NUMCULS  NUMPOS                SITE  NUMCULS  NUMPOS
    ASK    ASK     ASK   CONTAMINANT     ASK    ASK     ASK
                  RULE 006            RULE 007
             SITE      IDENT      SITE   IDENT   SUBTYPE
             ASK                  ASK
```

9

In the 1970s, symbolic AI was seen as the future of healthcare

By Gregory Freiherr

The Problems and Promises of Artificial Intelligence

In the past century, science has not only changed our conceptions about the world, it has changed itself. Driven by an explosion of information, specialties in science have sprung up, inevitably giving rise to subspecialties. But staying abreast of developments, even in narrowly specialized areas, is becoming increasingly difficult. One solution to management of this continuing flood of new information may be to create entities of intelligence.
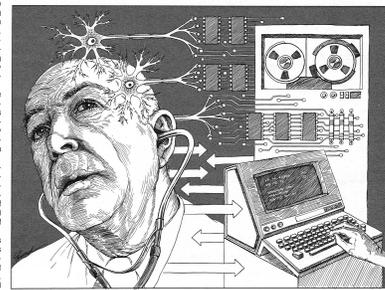
The proposed solution is the "intelligent machine," a computer that mimics the expert's reasoning power and can retain in retrievable memory much of the knowledge currently available to experts in a given specialty.

The branch of computer science that embodies such research is called artificial intelligence (AI). It is a multidisciplinary field, comprising teams of computer programmers and experts from intended areas of application.

In the long run, AI promises to amplify the skills of less specialized physicians or scientists and to fill the current gap in professional manpower. These programs may also lead to concise descriptions of the rules and processes used by experts in making decisions. Such descriptions would be of major use in education.

The network of the Stanford University Medical Experimental Computer for Artificial Intelligence in Medicine (SUMEX-AIM), which is supported by the Biotechnology Resources Program of the NIH Division of Research Resources, is a major force in research and development of AI. Many of the computer programs that now use AI techniques for biomedical

10

11

The excitement was real!

- But were systems made from a giant "soup" of rules really maintainable?
- Could semantic networks (knowledge graphs) support any kind of inference other than look-up?
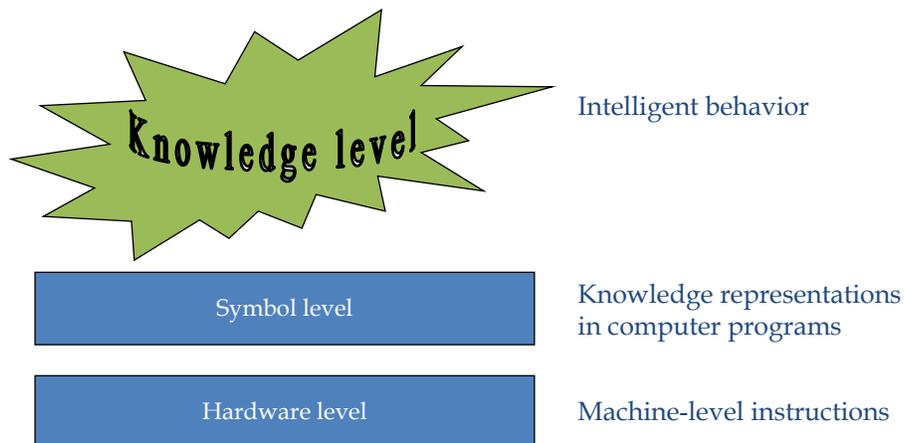- What knowledge representations might give us the "smartest" systems?

12

Allen Newell's AAAI Presidential Address (1980)

- We should stop bickering about how to represent knowledge
- What really matters is the knowledge that a system *has*, not how that knowledge is represented in the computer
- Knowledge is what an observer attributes to an agent to allow the observer to call that agent intelligent

13

A hierarchy of computer-system levels

Knowledge level — Intelligent behavior

Symbol level — Knowledge representations in computer programs

Hardware level — Machine-level instructions

14
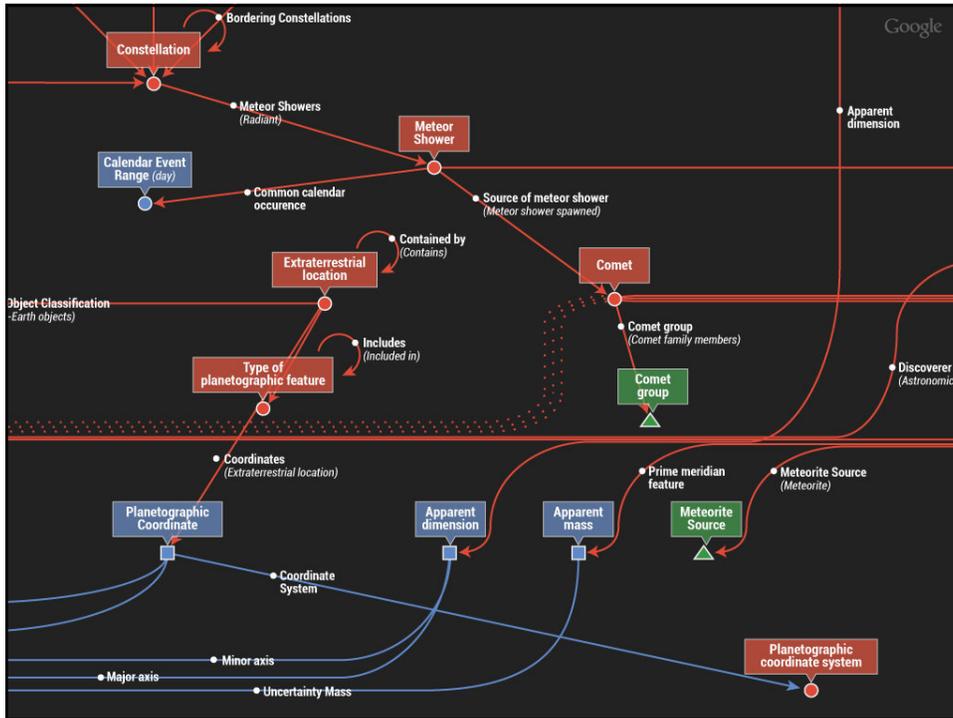
Knowledge is a competency for problem solving

- We can never actually "see" knowledge or write it down
- We can never really know what an agent knows
- We can only *attribute* knowledge to an agent if that agent
  - Appears to have goals
  - Appears to select actions to achieve its goals
  - Appears to select actions rationally

15

Knowledge-level analysis

- Offers ability to understand intelligent behavior in terms of
  - Goals
  - Actions
  - How actions are selected to achieve goals
- Requires a language for talking about goals and actions
- Makes underlying knowledge representation a secondary issue
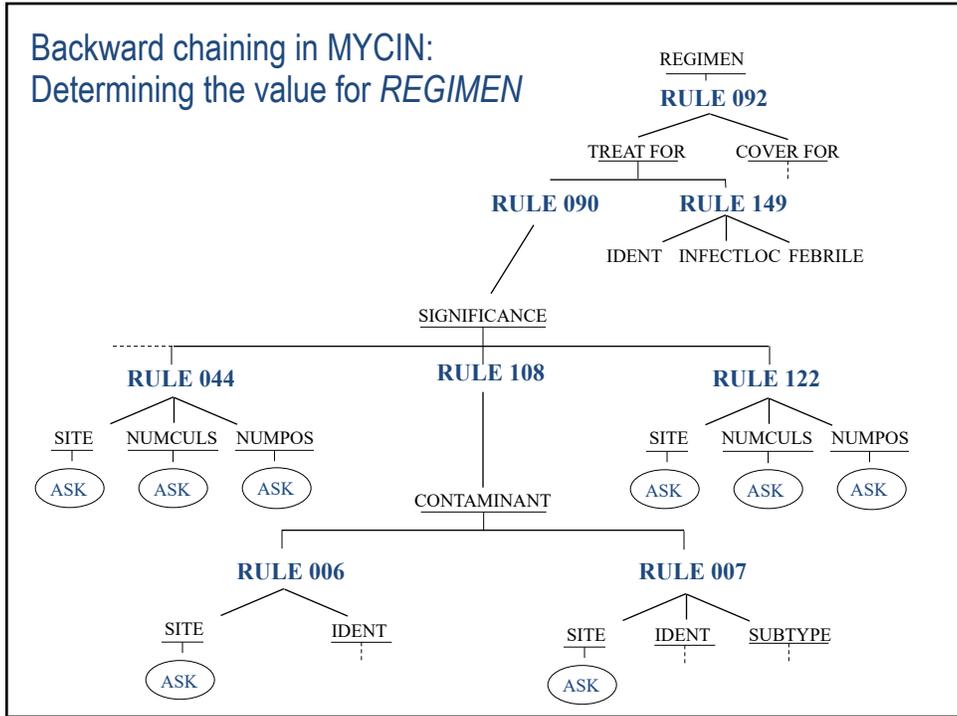
16

17

Knowledge is like music …

18

Slide 19

Backward chaining in MYCIN: Determining the value for *REGIMEN*

REGIMEN
RULE 092

TREAT FOR   COVER FOR

RULE 090   **RULE 149**

IDENT   INFECTLOC   FEBRILE

SIGNIFICANCE

RULE 044   **RULE 108**   **RULE 122**

SITE   NUMCULS   NUMPOS   SITE   NUMCULS   NUMPOS

ASK   ASK   ASK   ASK   ASK   ASK

CONTAMINANT

RULE 006   **RULE 007**

SITE   IDENT   SITE   IDENT   SUBTYPE

ASK   ASK

19

Slide 20

Heuristic classification in MYCIN (after Clancey)

Feature Abstraction   **Solution Refinement**

Compromised host   **Heuristic Match** →   Gram-negative infection

Immuno-suppressed   Alcoholic   Pseudo-monas   E. coli

Leukopenia

WBC < 2.5

20

Heuristic classification
in GRUNDY (after Clancey)

Feature Abstraction   **Solution Refinement**

Class of Person   **Heuristic Match** →   Class of book

Self-description and behavior

Books to recommend

21

Problem-Solving Methods (PSMs)

- Represent abstract procedures for solving stereotypical tasks
- Make explicit the roles in which domain knowledge is used during problem solving
- Provide reusable building blocks for composing abstract procedures

22

If you were to rebuild MYCIN for the 21st Century…

2. A problem-solving method that can use the ontology to identify likely pathogens and to recommend appropriate treatment
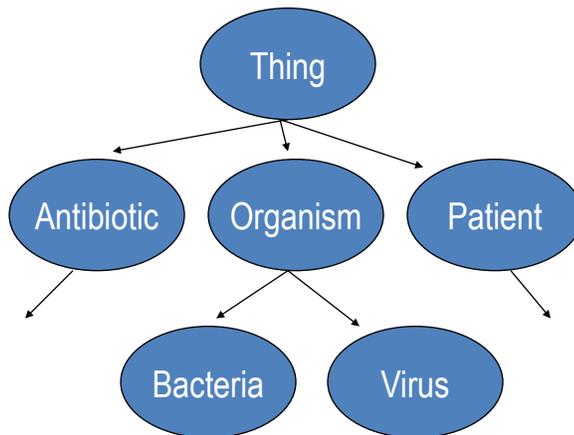
1. An ontology of infectious diseases

23

In the 1990s, the knowledge-based systems community had vast plans for libraries of reusable problem solvers that would perform

- Classification
- Fault diagnosis
- Constraint satisfaction
- Planning
- Design
- Scheduling
- Sequence alignment
- And a gazillion other abstract tasks

and to allow developers to model how agents
achieve their goals at the "knowledge level"

24

PSMs and task decomposition

- Tasks are solved by PSMs
- PSMs may entail subtasks, which themselves are solved by PSMs, which themselves may entail subtasks, which …

Task

Problem-Solving Method

Subtask   Subtask

PSM   PSM
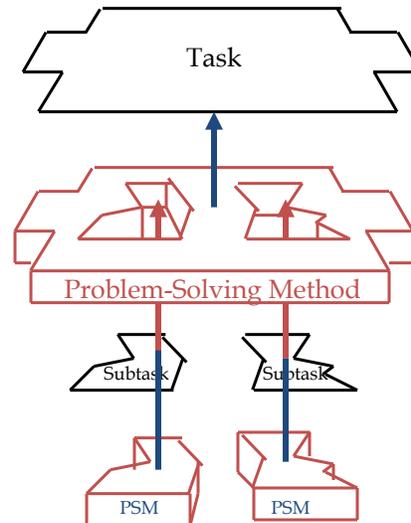
25

Modeling **tasks** was construed as developing a decomposition hierarchy of **subtasks** and the **problem-solving methods** that could address the subtasks

Do Laundry

Laundry method

Sort   Wash

Classification   Wash method
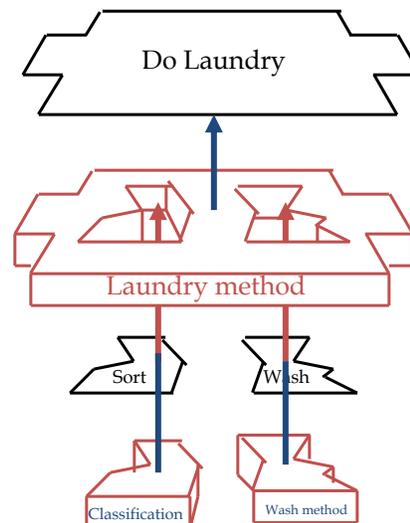
26

The Semantic Web was sold on the promise of problem solving at Web scale

- Personalized agents that could
  - Book flights
  - Shop for clothes
  - Manage our "smart homes"
  - Update our medical records
- A technology, like the Web itself, that would grow from the grassroots

27

But the world kept shifting

- The knowledge-based systems community seemed to drop everything and to jump onto the **Semantic Web** bandwagon
- The Semantic Web community began to downplay ontologies and Web services and to play up the idea of **linked data**
- The linked data that everyone cared about increasingly were represented in **knowledge graphs**
- We were back to a world where the focus was on representation; few people talked about problem solving, at least explicitly
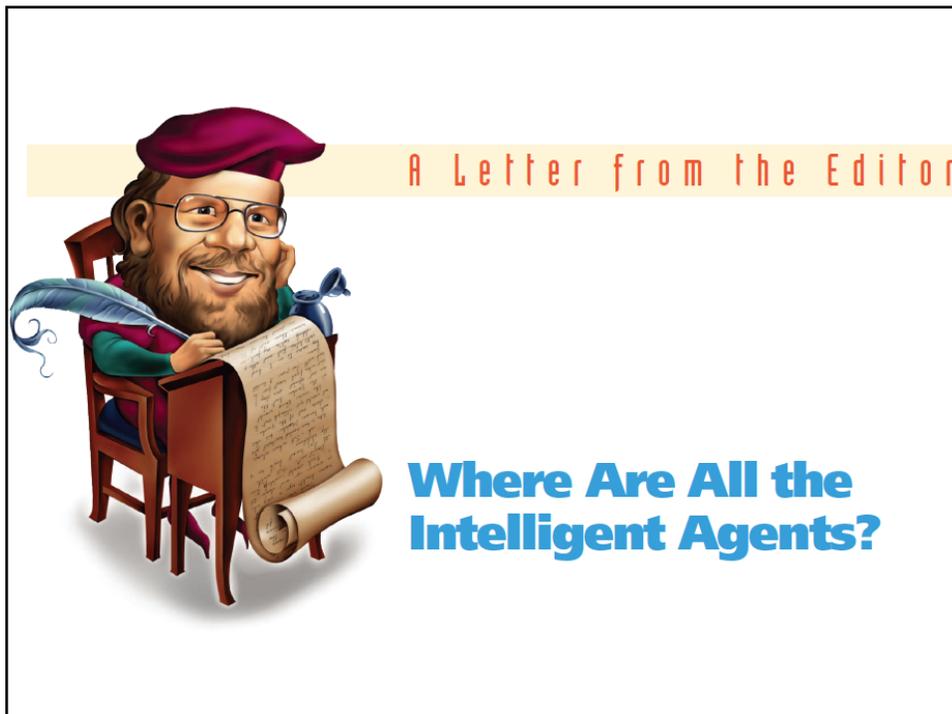
28

A Semantic Network (ca. 1970)

A directed graph where vertices are concepts and edges are relations
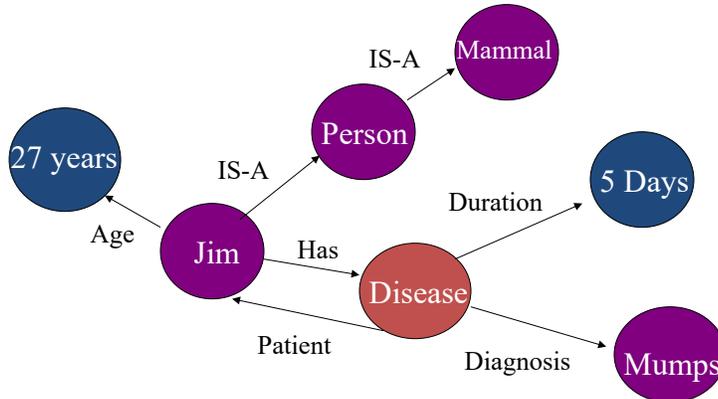
29

30

We can do wonderous things with knowledge graphs

- Performing indexing and search
- Driving chatbots
- Managing product catalogs
- Linking advertisers to consumers

But we are not yet taking on problem solving
the way we were twenty years ago.

31

# After 50 years, we've come full circle

- **Who'd a thunk:** We're back representing knowledge as graphs
- **The good news:** These graphs are enormous and much more interesting than our old semantic networks were
- **The disappointing news:** By themselves, these graphs still can't do anything
- **The opportunity:** We know how to apply processes to representations to generate intelligent behavior, and there are many behaviors that we haven't yet dreamed of modeling

32

What have we been talking about in this seminar?

- We all understand what a graph is

- But what do we mean by *knowledge*?

- And what does it mean to say that we have put knowledge in a graph?

33
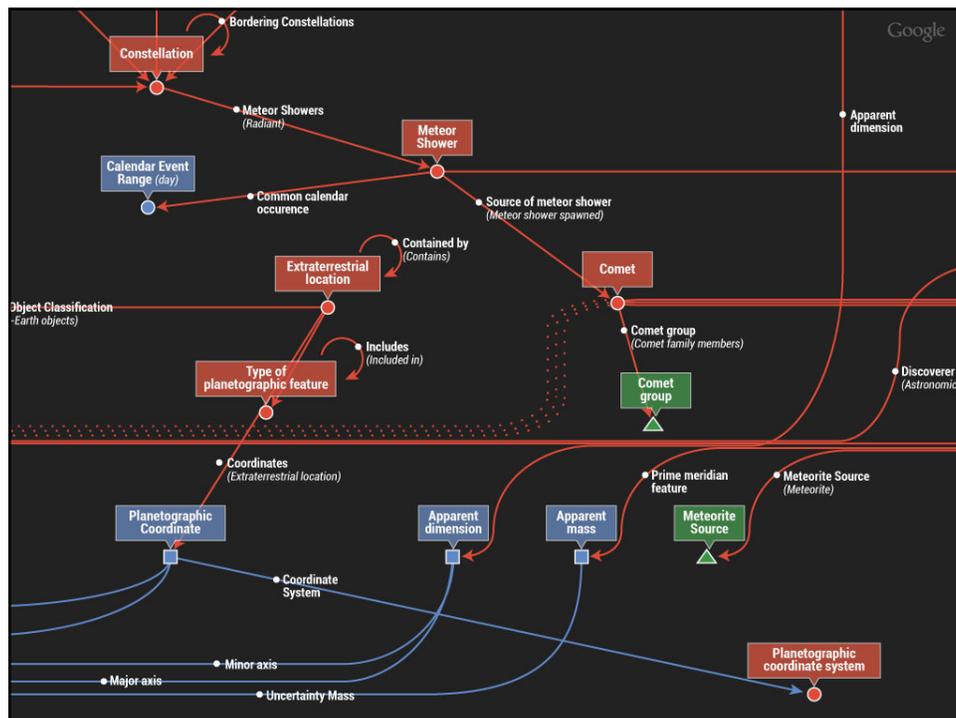
34

We want graphs that can make music …

35