Knowledge Engineering Basics for Accounting Professionals

By
Charles Hoffman, CPA
Charles.Hoffman@me.com

Objective of this Presentation

(Why this is important to professional accountants)

- **Financial analysis is already digital**, using business intelligence software (BI), electronic spreadsheets, etc.
- Need exists for a **digital version of the traditional general purpose financial report**
- **Professional accountants are knowledge workers** who rearrange abstract symbols (i.e. create financial reports)
- **Computers are tools**; machines which can be leveraged to rearrange abstract symbols (i.e. machine-readable information)
- XBRL-based financial reports are machine-readable structured information
- XBRL-based digital financial reports need to **work correctly** (i.e. cannot be a guessing game); **must convey meaning correctly**
- Professional accountants need to make digital financial reporting the way they need/desire it to work

Charles Hoffman, *Need for a Digital Alternative to General Purpose Financial Statement,* Retrieved February 24, 2016; http://xbrl.squarespace.com/journal/2015/10/28/need-for-digital-alternative-to-general-purpose-financial-st.html

Rearranging Abstract Symbols
(Why We Need Computers)

- In his book Saving Capitalism, Robert Reich describes three categories that all modern work/jobs fit into: (page 204)
  - **Routine production services** which entails repetitive tasks
  - **In-person services** where you physically have to be there because human touch was essential to the tasks
  - **Symbolic-analytic services** which includes problem solving, problem identification, and strategic thinking that go into the manipulation of symbols (data, words, oral and visual representations).
- In describing the third category, symbolic-analytic services, Mr. Reich elaborates:
  - In essence this work is to rearrange abstract symbols using a variety of analytic and creative tools - mathematical algorithms, legal arguments, financial gimmicks, scientific principles, powerful words and phrases, visual patterns, psychological insights, and other techniques for solving conceptual puzzles. Such manipulations improve efficiency-accomplishing tasks more accurately and quickly-or they better entertain, amuse, inform, or fascinate the human mind.
- Why this is interesting to me is the third category of work/jobs: symbolic-analytic services. Financial reporting, or at least many tasks related to financial reporting, fall into the symbolic-analytic service category.
- How many professional accountants think of their job as "rearranging abstract symbols using a variety of analytic and creative tools". Not many. Most professional accountants just do the work. Besides, what the heck is an "abstract symbol"?
- But that is what computers can do, rearrange abstract symbols, if information is represented using machine-readable metadata. Computers are good at repeating without variation

Robert B. Reich, *Saving Capitalism*, Alfred A. Knopf, page 204-206, (see a summary from The Work of Nations here http://www.oss.net/dynamaster/file_archive/040320/e8eb8748abfe77204a145d5fbcc892fb/OSS1993-01-37.pdf#page=6)

Understanding What Computers Cannot Do

- Understanding what computers cannot do is just as important as understanding what they can do; computers cannot replicate:
  - Intuition
  - Creativity
  - Innovation
  - Improvise
  - Exploration
  - Imagination
  - Judgement (such as making a tough decision from incomplete information)
  - Politics
  - Law
  - Unstructured problem solving
  - Non-routine tasks
  - Identifying and acquiring new relevant information
  - Compassion

Understanding Information

Data, Information, Knowledge, Wisdom

- **Data**: The basic compound for Intelligence is data -- measures and representations of the world around us, presented as external signals and picked up by various sensory instruments and organs. *Simplified: raw facts and numbers*.
- **Information**: Information is produced by assigning relevant meaning to data. *Simplified: information is data in context*.
- **Knowledge**: Knowledge is the subjective interpretation of information and approach to act upon the information in the mind of the perceiver. *Simplified: knowledge is the interpretation of information*.
- **Wisdom (or Intelligence)**: Intelligence or wisdom embodies awareness, insight, moral judgments, and principles to construct new knowledge and improve upon existing understanding. *Simplified: wisdom is the creation of new knowledge*.

DIKW Pyramid, Wikipedia, retrieved February 24, 2016; https://en.wikipedia.org/wiki/DIKW_Pyramid

Gene Bellinger, Durval Castro, Anthony Mills; *Data, Information, Knowledge, and Wisdom*; Retrieved February 24, 2016, http://www.systems-thinking.org/dikw/dikw.htm

Scenario from Financial Reporting

Two public companies, A and B, each have some knowledge about their financial position and financial condition. They must communicate their knowledge to an investor who is making investment decisions which will make use of the combined information so as to draw some conclusions. All three parties are using a common set of basic logical principles (facts known to be true, deductive reasoning, inductive reasoning, etc.) and common financial reporting standards (i.e. US GAAP, IFRS, etc.), so they should be able to communicate this information fully, so that any inferences which, say, the investor draws from public company A's input should also be derivable by public company A using basic logical principles and common financial reporting standards, and vice versa; and similarly for the investor and public company B.

Fundamental Challenge: Meaningful Exchange of *Information*

- The fundamental challenge is the **meaningful exchange of *information* between business systems**
- The ***only way*** a meaningful exchange of information can occur is the is the ***prior agreement*** as to
  - technical syntax rules,
  - business domain semantics rules, and
  - business domain workflow rules.

Video Introduction to HL7, Slide 4, Retrieved February 24, 2016, http://www.hl7.org/documentcenter/public_temp_D8292793-1C23-BA17-0C1CAB3A901C5581/training/IntroToHL7/player.html

Meaningful Exchange of *Information*

ISO TR 9007:1987 (“Helsinki principles”) state:

- Any meaningful exchange of utterances depends upon the prior existence of an agreed set of semantic and syntactic rules
- The recipients of the utterances must use only these rules to interpret the received utterances, if it is to mean the same as that which was meant by the utterer

Common Logic in Support of Metadata and Ontologies, Page 5, Retrieved June 24, 2016, http://cl.tamu.edu/docs/cl/Berlin_OpenForum_Delugach.pdf

Differentiating Technical Syntax and Domain Semantics

- Contrasting syntax and semantics
  - **Syntax** is how you say something
  - **Semantics** is the meaning behind what you said
- Technical syntax is not relevant to business professionals
- Domain semantics is important to business professionals

An Introduction to the Semantic Web, YouTube, Retrieved February 24, 2016; https://www.youtube.com/watch?v=OGg8A2zfWKg

Power of Agreement

- **Objective**: It is only through deliberate, methodical, rigorous and conscious collaboration, cooperation and coordination by the participants of the financial reporting supply chain that XBRL-based digital financial reporting will work safely, reliably, predictably, repeatedly, effectively, and efficiently.
- This objective will not be achieved by accident.
- Consider the definitions of arbitrary and standard:
  - **Arbitrary**: based on random choice or personal whim, rather than any reason or system; depending on individual discretion (as of a judge) and not fixed by law
  - **Standard**: used or accepted as normal; something established by authority, custom, or general consent as a model or example

Objective: What are we Trying to Achieve?

- Financial analysis is done using electronic spreadsheets and business intelligence (BI) software. Financial reports should be digital.
- What is the objective?
  - **To create a shared reality to achieve a specific purpose**: To arrive at a shared common enough view of "true and fair representation of financial information" such that most of our working purposes, so that reality does appear to be objective and stable. So that you can query information reliably, predictably, repeatedly, safely.
  - Meaningful information exchange that is reliable, repeatable, predictable, safe, cost effective, easy to use, robust, scalable, secure when necessary, auditable (track provenance) when necessary
- Prudence dictates that using the information contained in a digital financial report should not be a guessing game. Safe, reliable, repeatable, predictable, reuse of reported financial information using automated machine-based processes is obviously preferable to a guessing game.

William Kent updated by Steve Hoberman*, Data and Reality, Technics Publications*, (See this resource which has CHAPTER 12: Philosophy from the first version of this book, http://www.bkent.net/Doc/darxrp.htm)

Metadata

- **Metadata** is data that provides information about other data
- Metadata adds perspective and context
- Two types of metadata:
  - Descriptive
  - Structural
- Example of metadata is the card catalog of a library
- Classification system provides metadata

Wikipedia, *Metadata*, Retrieved March 1, 2016, https://en.wikipedia.org/wiki/Metadata

Classification Systems

- A **classification system** is a logical grouping of something based on some similarity or criteria. A classification system is a communications tool. A classification system structures information. A classification system can be informal or formal, more rigorously or less rigorously created, readable/usable by computers, or not. A classification system can be a controlled vocabulary.
- A **dictionary or list** is a classification system that has no hierarchy, it is simply a flat inventory of terms with no relations expressed between the terms. (But even a dictionary classifieds terms into noun, verb, adverb, etc.)
- A **taxonomy** is a classification system which tends to have one hierarchy into which some list of terms is categorized. Categories are basically sets. A taxonomy is a tree of categories of things with only one relation expressed so terms appear in only one location in a hierarchy of categories. A *taxonomist* creates concepts, creates coherent definitions for those concepts, and puts concepts into "buckets" or categories.
- An **ontology** is a classification system which tends to have more than one hierarchy into which terms are categorized. So an ontology can be thought of as a set of taxonomies. An ontology can express many different types of relations which includes traits/qualities of each term. An ontology is less like a tree and more like a graph (network theory). An *ontologist* identifies and establishes models explaining how things in a given taxonomy are related to one another, the kinds of relationships that exist, the rules of the model.

Network Theory, Wikipedia, retrieved February 24, 2016; https://en.wikipedia.org/wiki/Network_theory

What's the difference between Taxonomies and Ontologies?, retrieved February 24, 2016; http://www.ideaeng.com/taxonomies-ontologies-0602

Three Orders of Order

- **First order of order**. Putting books on shelves is an example the first order of order. (data)
- **Second order of order**. Creating a list of books on the shelves you have is an example of second order of order. This can be done on paper or it can be done in a database. (metadata)
- **Third order of order**. Adding even more information to information is an example of third order of order. Using the book example, classifying books by genre, best sellers, featured books, bargain books, books which one of your friends has read; basically there are countless ways to organize something. (more metadata)

David Wenberger, *Everything is Miscellaneous*, Holt Paperbacks, 2007; https://goo.gl/oj8mkf

Limitations of Classification Systems

- David Wenberger's book *Everything Is Miscellaneous* points out two important things about classification systems:
  - That every classification scheme ever devised inherently reflects the biases of those that constructed the classification system.
  - The role metadata plays in allowing you to create your own custom classification system so you can have the view of something that you want.

David Wenberger, *Everything is Miscellaneous*, Holt Paperbacks, 2007; https://goo.gl/oj8mkf

Business Rules are Expression of Knowledge, Metadata

- Another term used to describe relations is business rule.
- Business rules constrain and support facts.
- The Business Rules Manifesto does a good job of describing what a business rule is. Article 9; Of, By, and For Business People, Not IT People; points out the need for these business rules to be managed by business professionals:
  - 9.1. Rules should arise from knowledgeable business people.
  - 9.2. Business people should have tools available to help them formulate, validate, and manage rules.
  - 9.3. Business people should have tools available to help them verify business rules against each other for consistency.
- Business professionals are the ones who understand the problem domain. As such, business professionals are the ones who understand the business rules or relations between the things in their problem domain.

The Business Rules Manifesto, Business Rules Group, retrieved February 24, 2016; http://www.businessrulesgroup.org/brmanifesto.htm

Workflow Models

- Workflow the sequence of processes/tasks through which a piece of work passes from initiation to completion. https://en.wikipedia.org/wiki/Workflow
- There are two classes business workflow systems or models that business rules should be able to express rules that operate in both worlds.
  - **Process-centric workflows** generally use business rules at the workflow task level to manage workflow tasks.
  - **Data-centric workflows** generally use business rules within workflows to make decisions about individual items of data.
- This is not an "either/or" situation, but rather leveraging both workflow models in the design and execution of workflows is the way to go.
- By combining the two different workflow models, business rules can be undertaken at both the task level for automating different decisions and at the data level for implementing filters over the data. Business rules can also be used to define operational features of a workflow, such as what to do when a specific task fails.
  - Business Process Modeling (BPM) https://en.wikipedia.org/wiki/Business_process_modeling
  - XML Process Definition Language (XPDL) https://en.wikipedia.org/wiki/XPDL
  - Business Process Execution Language (BPEL) https://en.wikipedia.org/wiki/Business_Process_Execution_Language

Process-centric and Data-centric Business Workflow Systems, http://xbrl.squarespace.com/journal/2016/5/18/process-centric-and-data-centric-business-workflow-systems.html

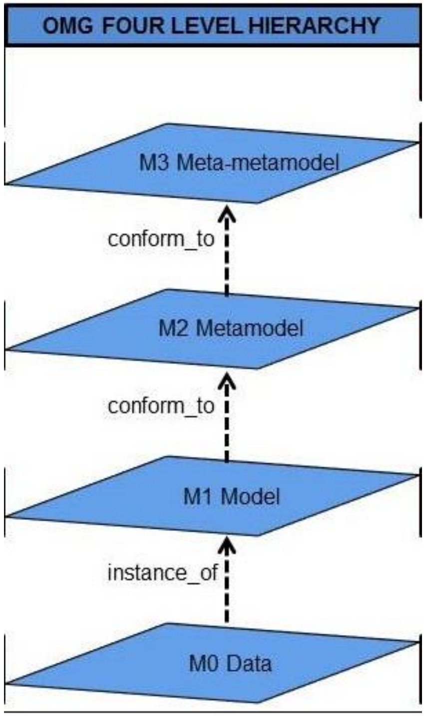
Need for Thick Metadata Layer of Knowledge

- What is not in dispute is the need for a "thick metadata layer" in order for the computer to be able to perform useful work.
- But what is sometimes disputed, it seems, is HOW to get that thick metadata layer.
- There are two basic approaches to getting this metadata layer:
  - **Have the computer figure out what the metadata is**: This approach uses artificial intelligence, machine learning, and other high-tech approaches to detecting patterns and figuring out the metadata.
  - **Tell the computer what the metadata is**: This approach leverages business domain experts and knowledge engineers to piece together the metadata so that the metadata becomes available.

Cognitive Computing And Semantic Technology: When Worlds Connect, Dataversity.net, retrieved February 24, 2016; http://www.dataversity.net/cognitive-computing-semantic-technology-worlds-connect/

Meta-Metamodel

- How do you share information *between models*?
- Meta-Metamodel
- The purpose of a meta-metamodel is to provide a framework to share information between systems effectively



Wikipedia, *Meta Object Facility*, Retrieved March 16, 2016, https://en.wikipedia.org/wiki/Meta-Object_Facility

Knowledgebase

- A knowledgebase is structure + data.
- For example, the XBRL taxonomies of public companies plus the XBRL-based financial reports of those companies in the SEC EDGAR system together is a knowledgebase.

Understanding Computers

Computers are Tools

- Computers are tools; Computers are machines
- Computer are unintelligent; they do not understand themselves, their programming, or the intended interpretation of the representations they contain and manipulate.
- Computers perform work which would otherwise need to be performed manually
- No magic is involved in the process of performing work

Andrew D. Spear, *Ontology for the Twenty First Century: An Introduction with Recommendations*; page 5, http://ifomis.uni-saarland.de/bfo/documents/manual.pdf#Page=5

Fundamental Strengths of Computers

- **Storage**: Computers can store tremendous amounts of information reliably and efficiently.
- **Retrieval**: Computers can retrieve tremendous amounts of information reliably and efficiently.
- **Processing**: Computers can process stored information reliably and efficiently, mechanically repeating the same process over and over.
- **Ubiquitous information distribution**: Computers can make information instantly accessible to individuals and more importantly other machine-based processes anywhere on the planet in real time via the internet, simultaneously to all individuals.

Andrew D. Spear, *Ontology for the Twenty First Century: An Introduction with Recommendations*; page 4, http://ifomis.uni-saarland.de/bfo/documents/manual.pdf#Page=4

Major Obstacles to Harnessing the Power of Computers

- **Business professional idiosyncrasies**: Different business professionals use *different terminologies* to refer to exactly the same thing.
- **Information technology idiosyncrasies**: Information technology professionals use *different technology options* , techniques , and formats to encode information and store exactly the same information.
- **Inconsistent domain understanding of and technology's limitations in expressing interconnections**: Information is not just a long list of facts, but rather these *facts are logically interconnected* and generally used within sets which can be dynamic and used one way by one business professional and some other way by another business professional or by the same business professional at some different point in time. *These relations are many times more detailed and complex than the typical computer database can handle*. *Business professionals sometimes do not understand that certain relations even exist.*
- **Computers are dumb beasts**: Computers don't understand themselves, the programs they run, or the information that they work with. Computers are "dumb beasts". What computers do can sometimes seem magical. But in reality, computers are only as smart as the metadata they are given to work with, the programs that humans create, and the data that exists in databases that the computers work with.

Andrew D. Spear, *Ontology for the Twenty First Century: An Introduction with Recommendations*; page 4, http://ifomis.uni-saarland.de/bfo/documents/manual.pdf#Page=4

Computers Need Structured Information

(Contrasting Unstructured and Structured Information)

- In the past financial reports were unstructured information, or at best information structured for presentation.
- Digital financial reports are structure for meaning using XBRL. See *How XBRL Works* Video.
- Unstructured can only be rearranged by humans as unstructured information is generally too complicated for machines to structure.
- Properly structured information can be rearranged using machines such as computers.

Charles Hoffman, *How XBRL Works,* YouTube, Retrieved February 24, 2016, https://www.youtube.com/watch?v=nATJBPOiTxM

Addressing “the 5 D’s” of Real-world Reasoning

Address “the 5 D’s” of real-world reasoning ⇒ desired improvements:

- **Diversity:** Existing/emerging kinds of ontologies and rules have heterogeneous knowledge rules. *Handle more heterogeneous systems*.
- **Distributed-ness:** of ownership/control of ontology/rule active knowledge bases. *Handle more source active knowledge bases*.
- **Disagreement:** Conflict (contradiction) will arise when merging knowledge. *Handle more conflicts.*
- **Dynamism:** Updates to knowledge occur frequently, overturning previous beliefs. *Handle higher rate of revisions*.
- **Delay:** Computational scalability is vital to achieve the promise of knowledge integration. *Achieve Polynomial-time (~databases).*

Benjamin Grosof and Mike Dean, *Objectives for Integrating Distributed SW Rules and Ontologies, Motivating SweetRules,* Retrieved June 22, 2016, http://www.mit.edu/~bgrosof/paps/talk-sweetrules-overview-2005-06-07.pdf#page=89

Logical Catastrophes Break Systems

- A logical catastrophe is a failure point. Logical catastrophes must be avoided. Systems should never have these failure points:
  - **Undecidability**: If a question cannot be resolved to a TRUE or FALSE answer; for example if the computer returns UNKNOWN then unpredictable results can be returned. Logic used by a computer must be decidable.
  - **Infinite loops**: If a computer somehow enters an infinite loop from which it cannot return because of a logic error or because the logic is too complex for the machine to work with; the machine will simply stop working or return nonsense.
  - **Unbounded system structures or pieces**: Systems need boundaries for them to work correctly. If a system does not have the proper boundaries, then a machine can become confused or not understand how to work with information that is provided.
  - **Unspecific or imprecise logic**: Confusing precise results with the capabilities of a computer to provide a statistically created result can cause problems. It is not expected that the business system at the level of describing the things in the system be able to support "fuzzy logic" or "probabilistic reasoning" or other such functionality.

Brainstorming Idea of Logical Catastrophes or Failure Points, Digital Financial Reporting Blog, retrieved February 24, 2016; http://xbrl.squarespace.com/journal/2015/7/25/brainstorming-idea-of-logical-catastrophes-or-failure-points.html

Understanding the critical importance of decidability

- In the **open world assumption** a statement cannot be assumed true on the basis of a failure to prove the statement. On a World Wide Web scale this is a useful assumption; however a consequence of this is that an inability to reach a conclusion (i.e. not decidable).
- In the **closed world assumption** the opposite stance is taken: a statement is true when its negation cannot be proven; a consequence of this is that it is always decidable. In other applications this is the most appropriate approach. So each application can choose to make the open world assumption or the closed world assumption based on its needs.
- **Because it is important that a conclusion** as to the correct mechanics of a financial report is required because consistent and correct mechanics are necessary to making effective use of the information contained within a financial report; the system used to process a financial report must make the closed world assumption.
- **Decidability** means that a conclusion can be reached.

Identity

- Computers need a way to grab onto information, or “handles” which they can work with.
- Things may or may not be unique. What does “us-gaap:StatementTable” mean?
  - **Isomorphic**: Has one meaning.
  - **Polymorphic**: Has more than one meaning.
- Identifiers can change

Charles Hoffman, *Understanding Why SEC XBRL Financial Information is Hard to Use and How to use Prototype Theory to Solve that Problem*, http://goo.gl/sYP2O7

Difference Between Notion/Idea/Phenomenon, Name, and Preferred Label

- Important to understand the difference/distinction between:
  - Notion, idea, phenomenon: something that exists in reality
  - Name: identifies some notion/idea/phenomenon (are standard)
  - Preferred label: alternative ways used to refer to name (tend to be arbitrary)

Prototype Theory

- Fundamentally there are two perspectives to *understanding what something is*:
  - Aristotle's perspective was "A thing is a member of a category if it satisfies the definition of the category."
  - The second perspective, **prototype theory**, is that we can know what something is even if it can't be clearly defined and even if its boundaries cannot be sharply drawn; concepts can be clear without having clear definitions if they're organized around undisputed examples, or prototypes.
- Example: one can understand that something is a "chair" by understanding as many properties as possible about the thing you are looking at, looking at the properties of a chair as defined by a prototype (the undisputed example), and then predicting whether the thing you are looking at is a "chair" by comparing the properties you are looking at with the properties of a chair.
- By contrast, the definitional view "draws sharp lines" whereas the prototype view works because "things can be sort of, kind of in a category. Prototype theory relies on our implicit understanding and does not assume that we can even make that understanding explicitly.
- In addition to prototype theory, there is exemplar theory and multiple-prototype theory.

Eleanor Rosch, *Principles of Categorization*, http://commonweb.unifr.ch/artsdean/pub/gestens/f/as/files/4610/9778_083247.pdf

Wikipedia, *Prototype Theory*, Retrieved February 24, https://en.wikipedia.org/wiki/Prototype_theory

Taxonomic Keys

- A identification key or taxonomic key is a printed or computer-aided device used for identifying some entity that is unknown. Keys are constructed so that the user is presented with a series of choices about the characteristics of the unknown thing; by making the correct choice at each step of the key, the user is ultimately led to the identity of the thing. Taxonomic keys are also helpful in classifying things into a standard taxonomy consistently.
- There are two types of keys:
  - **Single-access keys**: A single-access key (dichotomous key, sequential key, analytical key, or pathway key) is an identification key where the sequence and structure of identification steps is fixed by the author of the key.
  - **Multi-access keys**: A multi-access key enables the user to freely choose the set and characteristics that are convenient to evaluate for the item to be identified.
- Single-access keys and multi-access keys serve the same purpose. Each approach has its advantages and disadvantages.
- One advantage of multi-access keys is that users can enter or select information about an unidentified thing in any order, allowing the computer to interactively rule out possible identifications of the entity and present the user with additional helpful information and guidance on what information to enter next. A disadvantage of multi-access keys is that you have to understand a certain amount about a domain to use them; the more you understand about a domain the more useful multi-access keys can be.
- One advantage of single-access keys is that if you don't understand the domain or don't understand enough the single-access keys can serve as bread crumbs that provide a path to the answer you are looking for.

Understanding Taxonomic Keys, http://xbrl.squarespace.com/journal/2016/5/13/understanding-taxonomic-keys.html

Expert System

- Expert systems are computer programs that are built to mimic or simulate or emulate human behavior and knowledge; expert systems are computer application that performs a task that would otherwise be performed by a human expert.
- Expert systems solve problems by reasoning about knowledge represented in machine-readable form as "IF…THEN" rules that the machine simply follows.
- Computers really are not thinking, they are only mimicking or simulating or emulating human though by following a clearly laid out set of machine-readable instructions to perform some task.
- Frank Puppe explains in his book that there are three general categories of expert systems:
  - **Classification or diagnosis type**: helps users of the system select from a set of given alternatives.
  - **Construction type**: helps users of the system assemble something from given primitive components.
  - **Simulation type**: helps users of the system understand how some model reacts to certain inputs.

Frank Puppe, *Systematic Introduction to Expert Systems, Knowledge Representations and Problem-Solving Methods*, page 11 (Note that you can read Parts I and II on Google Books here, https://books.google.com/books?id=_kKqCAAAQBAJ)

Components of an Expert System

A software based expert system has four primary components:

- **Database of facts**: A database of facts is a set of observations about some current situation or instance. The database of facts is "flexible" in that they apply to the current situation. The database of facts is machine-readable. An XBRL instance is a database of facts.
- **Knowledge base**: A knowledge base is a set of universally applicable rules created based on experience and knowledge of the practices of the best domain experts generally articulated in the form of IF…THEN statements or a form that can be converted to IF...THEN form. A knowledge base is "fixed" in that its rules are universally relevant to all situations covered by the knowledge base. Not all rules are relevant to every situation. But where a rule is applicable it is universally applicable. All knowledge base information is machine-readable. An XBRL taxonomy is a knowledge base.
- **Rules processor/inference engine**: A rules processor/inference engine takes existing information in the knowledge base and the database of facts and uses that information to reach conclusions or take actions. The rules processor/inference engine is the machine that processes the information.
- **Explanation mechanism**: The explanation mechanism explains and justifies how a conclusion or conclusions are reached. It walks you through which facts and which rules were used to reach a conclusion. The explanation mechanism is the results of processing the information using the rules processor/inference engine and justifies why the conclusion was reached.

Components of an Expert System, http://xbrl.squarespace.com/journal/2016/5/24/understanding-the-components-of-an-expert-system.html

Benefits of an Expert System

Benefits from the use of expert systems include:

- **Automation**: elimination of routine, boring, repetitive, mundane, mechanical tasks that can be automate
- **Consistency**: computers are good at performing repetitive, mechanical tasks whereas humans are not; computers do not make mistakes and are good at repeating exactly the same thing each time
- **Diligence and tenacity**: computers excel at paying attention to detail; they never get bored or overwhelmed and they are always available and will keep doing their job until the task is complete with the same attention to detail
- **Reduced down-time**: computer based expert systems are tireless and do not get distracted
- **Availability**: such computer based expert systems are always available simultaneously in multiple places at one time; you get quick response times and can replace absent or scarce experts
- **Training**: the best practices of the best practitioners can be available to those that are new to and learning about a domain of knowledge
- **Longevity and persistence**: computer based expert systems do not change jobs or retire so knowledge gathered by an organization can remain within that organization
- **Productivity**: computer based expert systems are cheaper that hiring experts and costs can be reduced a the same time that quality increases resulting in increased productivity
- **Multiple opinions**: Systems can integrate the view of multiple experts within a system and choose between the preferred view of multiple expert opinions in the same system
- **Objectivity**: computers apply the same inductive and deductive logic consistently; emotion and personal preferences can be eliminated where they should be eliminated

Benefits of an Expert System, http://xbrl.squarespace.com/journal/2016/5/24/understanding-the-components-of-an-expert-system.html

Artificial Intelligence

- **Artificial intelligence** is the automation of activities that we associate with human thinking and activities such as decision making, problem solving, learning and so on.
- **Expert systems** is a branch of artificial intelligence. Expert systems are computer programs that are built to mimic human behavior and knowledge.

Introduction to Artificial Intelligence Terminology, Retrieved July 24, 2016; http://xbrl.squarespace.com/journal/2016/7/21/introduction-to-artificial-intelligence-terminology.html

Understanding the Components of an Expert System, Retrieved July 24, 2016; http://xbrl.squarespace.com/journal/2016/5/24/understanding-the-components-of-an-expert-system.html

Intelligent Agent

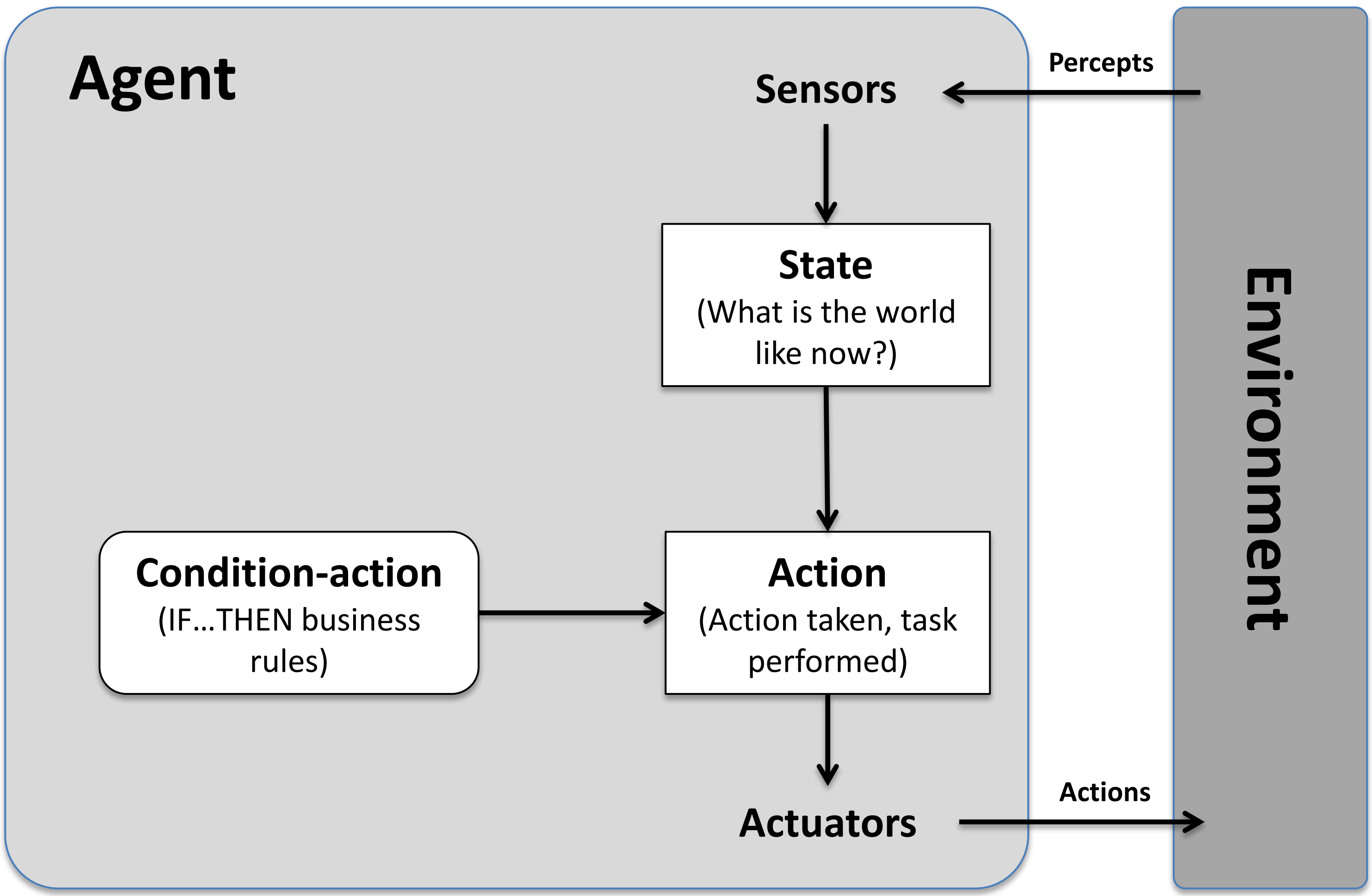
- An ***intelligent agent*** is software that assists people and acts on their behalf. Intelligent agents work by allowing people to:
  - **delegate work** that they could have done to the agent software.
  - **perform repetitive tasks**,
  - **remember things** you forgot,
  - intelligently find, filter and **summarize complex information**,
  - **customize** information to your preferences,
  - **learn from you** and even make recommendations to you.

Introduction to Intelligent Agents for Business Professionals, Retrieved July 24, 2016;
http://xbrlsite.azurewebsites.net/2016/Library/02_IntroducingIntelligentAgents.pdf
Hanh Tran & Thaovy Tran, *Intelligent Agent*, Retrieved July 24, 2016;
http://groups.engin.umd.umich.edu/CIS/course.des/cis479/projects/agent/Intelligent_agent.html

Agent

- An agent is an entity capable of **sensing** the **state** of its **environment** and **acting** upon it based on a set of specified **rules**. An agent performs specific tasks on behalf of another. In the case of software, an agent is a software program.
- The main difference between a software agent and an ordinary program is that a software agent is **autonomous**; that is, it must operate without direct intervention of humans or others.

Simple Reflex Agent



Wikipedia, *Intelligent Agent,* Retrieved July 24, 2016; https://en.wikipedia.org/wiki/Intelligent_agent

Understanding Knowledge Engineering Tools and Techniques

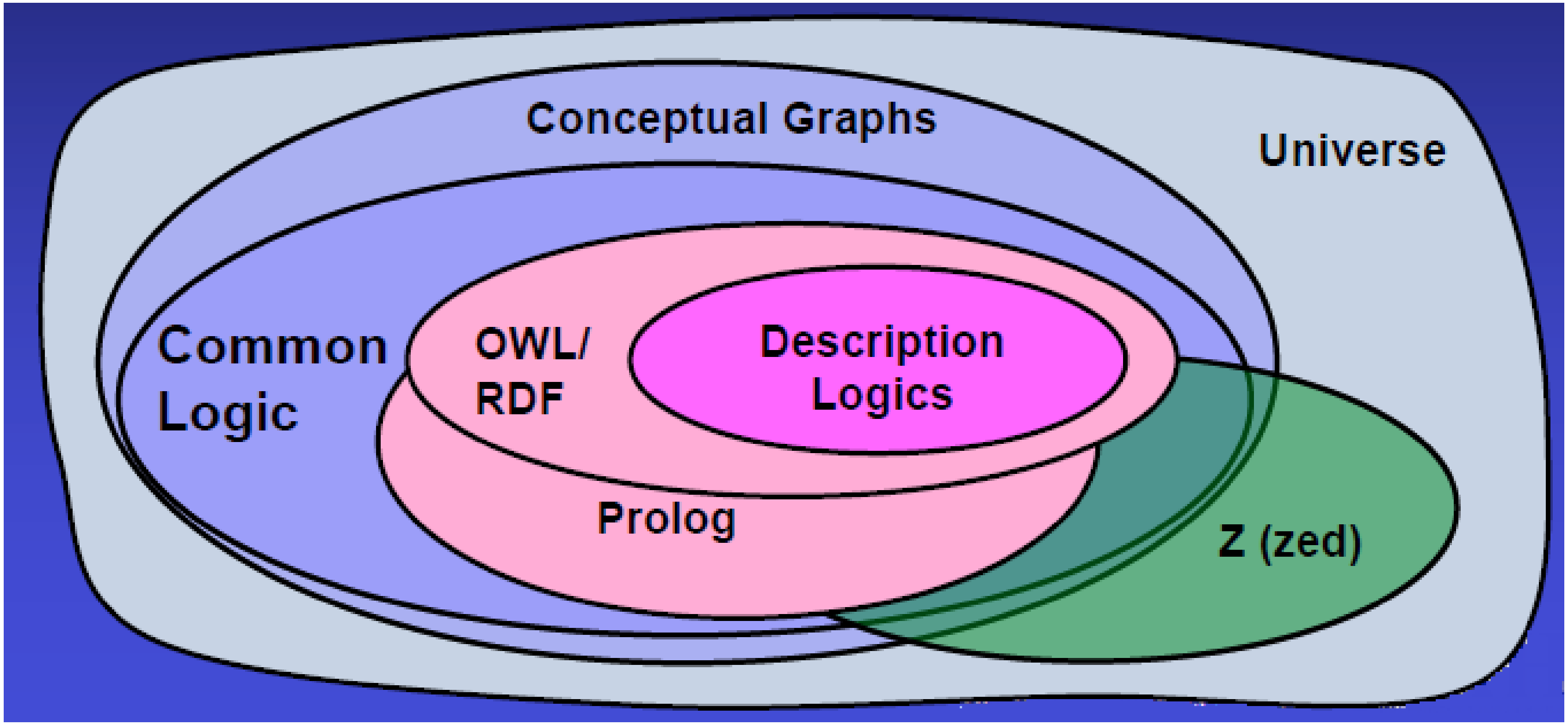
Setting the Right Expectations

- Business professionals need to understand what computers can and cannot do. Computers cannot perform magic . Computers fundamentally follow the rules of mathematics which follow the rules of formal logic. It really is that straight forward. Computers cannot effectively work with information such as the following:
  – fuzzy expressions - "It **often** rains in autumn."
  – non-monotonicity - "Birds fly, penguin is a bird, but penguin does not fly."
  – propositional attitudes - "Eve **thinks** that 2 is not a prime number." (It is true that she thinks it, but what she thinks is not true.)
  – modal logic
    - possibility and necessity - "It is **possible** that it will rain today."
    - epistemic modalities - "Eve **knows** that 2 is a prime number."
    - temporal logic - "I am **always** hungry."
    - deontic logic - "You **must** do this."

Limitations of First-order logic expressiveness, http://dior.ics.muni.cz/~makub/owl/

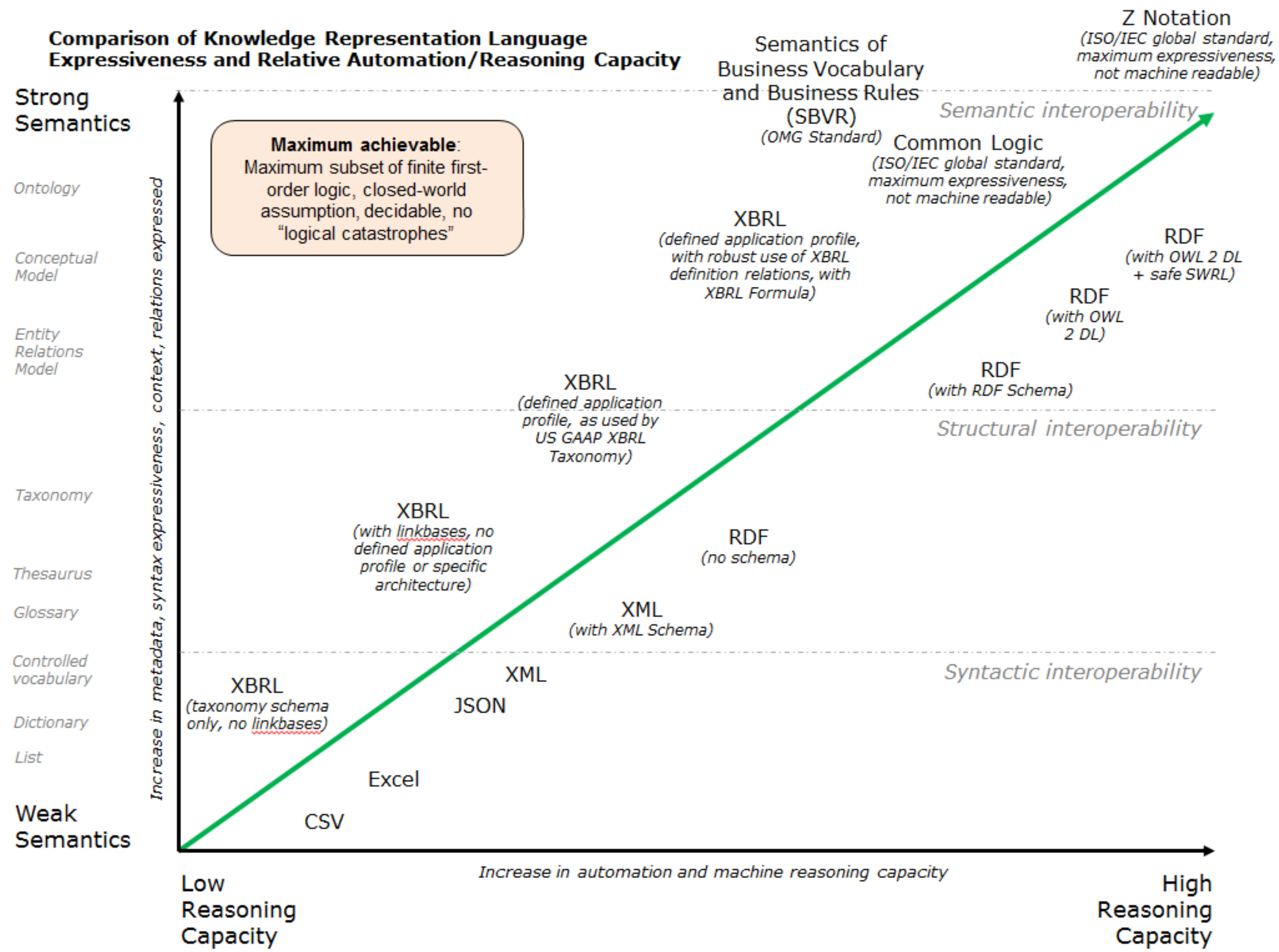
Comparing Expressiveness

Expressiveness is the set of things that can possibly be expressed by some language. For example, first order logic does not express modalities such as possibility.



Common Logic in Support of Metadata and Ontologies, Page 2, Retrieved June 24, 2016, http://cl.tamu.edu/docs/cl/Berlin_OpenForum_Delugach.pdf

Relation Between Expressiveness and Reasoning Capacity



Inspired by similar comparisons from *An Intrepid Guide to Ontologies* http://www.mkbergman.com/date/2007/05/16/ and *Semantics Overview* http://prezi.com/prwsxj8po3ln/semantics-overview/

Ontologies are Tools

- The term ontology has been used in philosophy for thousands of years going back to the father of formal logic, Aristotle (400 B.C.). Ontology is defined as the study of the things and the relations between things that exist in reality. The goal of philosophical ontology is to provide deliberate, clear, coherent and rigorously worked out accounts of the basic structures found in reality.

Ontology

- An ontology is a salient collection of the classes and subclasses of a problem domain or area of concern. An ontology accurately represents reality. The goal of an ontology is to provide a deliberate, rigorously and methodically worked out, description of the important things and relations between things which is clear, consistent, logically coherent, and unambiguous.
- Ontologies identify and describe sets of basic categories of things and universal similarities that these sets have. Universals can be instantiated by more than one object at more than one time. Particulars are non-repeatable and can exist in only one place at any one given time. So, things can be universals or things can be particulars. Relations can exist between universals and particulars:
  - **Universal to universal relations**, “is-a” or “is a subtype of” type relations
  - **Universal to particular relations**, “has-property” type relations
  - **Particular to particular relations**, “part-of” type relations

Intrepid Guide to Ontologies, Retrieved February 24, 2016; http://www.mkbergman.com/date/2007/05/16/

Understanding What Ontologies Do

- **Ontologies overcome the four major obstacles of getting a computer system to perform work discussed previously**. Remember the goal: reliable, repeatable, predictable, safe, cost effective, easy to use, robust, scalable, secure when necessary, auditable (track provenance) when necessary.
- **Ontologies are formal specifications**. Formal specifications are precise, concise and unambiguous. Formal specifications are communications tools. Because ontologies use machine-checkable notation, a wide variety of automated checks can be applied. The disciplined approach of using formal specifications means that subtle errors and oversights will be detected and corrected.
- **Ontologies both describe the information being worked with and verify information consistency against that description to avoid information quality problems or inconsistencies**. When creating information it is important to verify that what has been created is consistent with the expected description. When consuming information it is important to understand that the information being consumed is consistent with the expected description. Remember: nonsense-in-nonsense-out.

Robert Arp, Barry Smith and Andrew D. Spear, *Building Ontologies with Basic Formal Ontology,* MIT Press 2015; https://mitpress.mit.edu/building-ontologies

Conceptual Information Model

- A **conceptual model** is an abstraction of things in the real world which is used to help people understand the subject or domain the model represents. A conceptual model is composed of types of concepts, concepts, and rules which describe relations between types of concepts.
- An **ontology** is a type of conceptual model. Other closely associated terms are logical model, entity-relationship model.
- A **theory** is a prescriptive or normative statement which makes up a body of knowledge about what ought to be. A theory provides goals, norms, and standards. To theorize is to develop this body of knowledge.
- A conceptual model, an ontology, and a theory all tend to serve the same general purpose which is to describe a domain of knowledge.

Wikipedia*, Conceptual Model,* Retrieved August 14, 2016; https://en.wikipedia.org/wiki/Conceptual_model

Multidimensional Model

- Models help communication and provide a framework for understanding
- Everyone works with multidimensional information every day
- Multidimensional model overview
  - Fact
  - Characteristic
  - Fact Value
  - Fact Table
  - Relation
  - Grain

Representing Domain Knowledge

Difference Between “Simple” and “Simplistic”

- Anyone can create something that is sophisticated and complex.
- It is much harder to create something that is sophisticated and simple.
- Simple is not the same thing as simplistic. "Simple" is not about doing simple things. Simple is the ultimate sophistication. Simple is elegant.
- Simplicity is “dumbing down” a problem to make the problem easier to solve. That is not what simple is about.
- Simple is about beating down complexity in order to make something simple and elegant; to make sophisticated things *simple to use* rather than *complex to use*.

Differentiating Important and Unimportant

- The following terms help one understand the difference between an important nuance and an unimportant negligible difference.
  - **Nuance**: a subtle but important difference in or shade of meaning, expression, or sound; a subtle but important distinction or variation
  - **Subtle**: so delicate or precise as to be difficult to analyze or describe but important; hard to notice or see but important; not obvious but important
  - **Negligible**: so small or unimportant as to be not worth considering; insignificant; so small or unimportant or of so little consequence as to warrant little or no attention
- *Business professionals can best differentiate important nuances from unimportant negligible differences*. They do not do it perfectly and the only real way to make sure things are right is testing and experimentation at times.
- **Ontologies are about getting the salient aspects of a problem domain right**. One needs to take a pragmatic view of the world because it is impossible to describe every single aspect of the world. Ontologies only need to represent the important things. An ontology is therefore more like a “wireframe” or a “substrate”.

Important Knowledge Engineering Terms

- **Thing**: A *thing* is something that exists in the real world, in the problem domain, in the area of concern. A thing is just a class that all classes and individuals of the problem domain must belong to. All classes are subclasses of thing. Every individual must be of some class. Every class is a thing. Therefore since all classes are subclasses of thing; then all individuals are likewise ultimately a thing. “Nothing” is the opposite of thing.
- **Class**: A *class* is a set of individuals that have one or more similar distinguishing features in common. Classes are universals. For example person is the class consisting of all persons of which Bill Gates III is a member. Each problem domain can be captured in terms of a family of classes, together with a set of relations. The most important relation is the subclass relation (also called is-a) which organizes the classes in a taxonomic tree. Other key types of relations are whole-part and has-part.
- **Individual**: An *individual* is some specific item that exists in reality. Individuals are particulars. For example, a specific person such as Bill Gates III, a specific report such as Fiscal year 2014 financial statement, a specific economic entity such as Microsoft Corp. An individual exists only once.
- **Property**: (universal to particular, has property) A *property* is a trait, quality, feature, attribute of an individual, for example the property of being male of a person, of being filed of a report, and so on.
- **Relations between individuals**: (particular to particular, part of) one individual can be related to another individual, as when Bill is brother-of Dave, Bill is owner-of the building at 1835 73rd Ave NE, Medina, and so on.
- **Relations between classes**: (universal to universal, is subtype of) when every member of a certain class stands in a certain relation to some member of another class, then the relation is universal and we can formulate this as a relation between classes. So for example because every brother is identical to some male person, we can assert this as a relation between the classes brother and male person to the effect that brother is-a male person – in other words the class brother is included as a subclass in the class male person. If every financial report has some statement as part, then we can assert financial report has-part statement. Relations between classes are universal and apply to every member of that class.

Robert Arp, Barry Smith and Andrew D. Spear, *Building Ontologies with Basic Formal Ontology,* MIT Press 2015; https://mitpress.mit.edu/building-ontologies

Pitfalls of Knowledge Engineering

- Logical errors
- One rigid reality (i.e. *Everything is Miscellaneous*)
- Overly complicated representation (i.e. *simplicity*)
- Blind trust of domain experts
- Misuse of highly-expressive languages

Pitfalls are Avoidable

- All pitfalls are avoidable with conscious attention to detail, skillful execution, and rigorous testing

Double Entry Accounting

- **Single entry bookkeeping** is how 'everyone' would do accounting.
- **Double entry bookkeeping** adds an additional important property to the accounting system; that of a clear strategy to identify errors and to remove them. Even better, it has a side effect of clearly firewalling errors as either accident or fraud.
- This then leads to an audit strategy.
- Double entry bookkeeping, invented by Luca Pacioli, is one of the greatest discoveries of commerce, and its significance is difficult to overstate.

Ian Grigg, *Triple Entry Accounting,* http://iang.org/papers/triple_entry.html

Understanding Systems

A Digital Financial Report is a System

- Financial reporting is a system
- A financial report is part of the financial reporting system
- A digital financial report *itself* is a mechanical system (objective)
- Deciding *what should go into* a financial report is subjective (requires professional judgement)

Dr. W. Edwards Deming, YouTube, *A Theory of Systems for Educators and Managers*, Retrieved February 24, 2016; https://www.youtube.com/watch?v=2MJ3lGJ4OFo

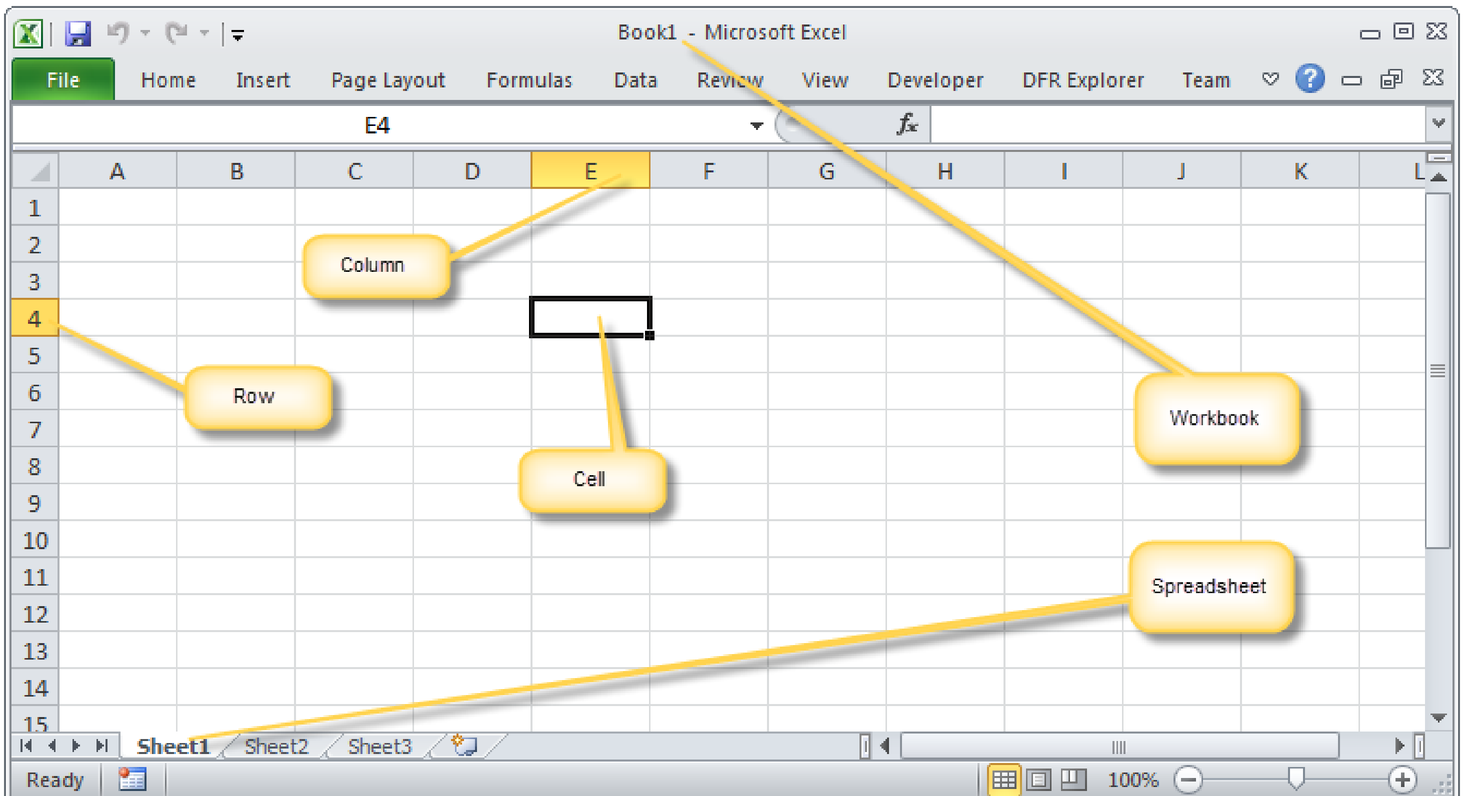
Describing Systems Formally

- Aristotle is said to be the father of formal logic. Logic is a discipline of philosophy. Logic is the study of correct reasoning.
- The notation of what we call elementary school arithmetic took centuries to develop. But today we take mathematics for granted.
- Formal logic is the basis for mathematics. Mathematics is a formal system.
- A formal system is defined as any well-defined system of abstract thought based on the model of mathematics. Of particular importance in describing systems is set theory. Basically, formal systems can be explained and proven to work or show system flaws and inconsistencies using the language of mathematics. Every formal system has some sort of formal language that explains that system. Every formal system can be tested to see if it works using a formal proof.
- A theory is a tool for understanding, explaining, and making predictions about a system. A theory describes absolutes. A theory describes the principles by which a system operates. A theory can be right or a theory can be wrong; but a theory has one intent: to discover the essence of some system.
- A theory is consistent if its theorems will never contradict each other. Inconsistent theories cannot have any model, as the same statement cannot be true and false on the same system. But a consistent theory forms a conceptual model which one can use to understand or describe the system. A conceptual model or framework helps to make conceptual distinctions and organize ideas.
- Formal logic was consciously broken into two groups: first-order logic and higher-order logic. There is a reason for this. Systems based on first-order logic can be proven to be **sound** (all provable theory statements are true in all models) and **complete** (all theory statements which are true in all models are provable using proof theory).
- Basically, higher-order logics are less well-behaved than those of first-order logic. They are less predictable and therefore less reliable and they are significantly harder to implement using computers.
- That is why computer systems are generally based on first-order logic.
- Business rules that are based on first-order logic prevent anarchy.
- This is all well understood by good software engineers. A finite-state machine is a model for designing computer systems that should never break. Finite-state machines are Turing machines. These machines tend to not "freeze" or "fail" or fail to return an answer. The Chomsky Hierarchy categories languages into groups that provide different levels of reliability. Type-0, type-1, type-2, and type-3 languages all are Turing machines.
- How does all this relate to professional accountants?
- First, bear in mind that you don't need to be a nuclear engineer to plug an appliance which uses electricity. You don't need to be an automotive engineer to understand how to start and drive a car. You don't need to be a hydraulics engineer to take a drink of water from a faucet. But you do want good engineers to build those systems so the systems don't break.
- A digital financial report is a formal system. A system such as the digital financial report needs to be described precisely so that professional accountants understand how the system works so that the system can be used effectively and so the system works how the system was intended to work. Z Notation is an ISO/IEC standard for describing systems precisely. But Z Notation is not machine-readable.
- Common Logic (CL), also an ISO/IEC standard, is a framework for a family of logic languages, based on first-order logic, intended to facilitate the exchange and transmission of knowledge in computer-based systems. Common Logic is machine-readable.
- Semantics of Business Vocabulary and Business Rules (SBVR) is an OMG standard that was designed and built to be logically equivalent to Common Logic
- Common Logic is about being practical, something business professionals generally tend to like. Common logic is a conscious compromise in order to achieve reliability, predictability, and safety. Common Logic is a "sweet spot" that achieves high expressivity but consciously gives up certain specific things that lead to catastrophic results that cause systems to potentially break making a system unsound; so that a system will be sound. Common Logic establishes boundaries, allowing creators of systems to "stay within the lines" and if you do, you get a maximum amount of expressiveness with the minimum risk of catastrophic system failure. Thus, you get a more reliable, dependable system.
- Deliberate, rigorous, conscious, skillful execution is preferable to haphazard, negligent, unconscious, inept execution if you want to be sure something works.
- Another way to describe a system is to create a theory and then prove the theory. *Financial Report Semantics and Dynamics Theory* is such a theory and proof. Digital financial reports can be tested for consistency against a documented theory. Consistencies with the theory help prove the theory. Inconsistencies with the theory can be detected in the system and the reason for the inconsistency determined.
- If you understand how formal logic works all this makes sense. If you don't understand formal logic, please watch this video which is a crash course in formal logic.
- Systems need to work properly to be useful. That is what engineering is all about, making things work properly so that they are useful. Luck and magic have nothing to do with making systems work. Conscious, deliberate, and skillful attention to the details of a system contribute to making a system work.
- It would be very hard to argue that XBRL-based financial reports to the SEC are a resounding failure. Likewise, it would be hard to argue that XBRL-based financial reports to the SEC are an overwhelming success given the quality problems that exist with those digital financial reports today. What one can do given the right tools is to gather empirical evidence about how the system is working and compare that evidence to how one thinks the system should work.

Charles Hoffman, *Describing Systems Formally*, Retrieved February 24, 2016; http://xbrl.squarespace.com/journal/2016/2/11/describing-systems-formally.html

Models Help Communication and Understanding

- Electronic spreadsheet model
  - Workbook, spreadsheet, column, row, cell



Difference between a Metaphor, a Model, and a Theory

- A **metaphor** describes something less understandable by relating it to something more understandable.
- A **model** is a specimen that exemplifies the ideal qualities of something. Models tend to simplify. There tend to always be gaps between models and reality. Models are analogies; they tend to describe one thing relative to something else. Models need a defense or an explanation.
- A **theory** describes absolutes. Theories are the real thing. A theory describes the object of its focus. A theory does not simplify. Theories are irreducible, the foundation on which new metaphors can be built. A successful theory can become a fact. A theory describes the world and tries to describe the principles by which the world operates. A theory can be right or wrong, but it is characteristic by its intent: the discovery of essence.

Jeff Hawkins, Ted Talk, *How brain science will change computing*, Retrieved February 24, 2016; https://www.youtube.com/watch?v=G6CVj5IQkzk

Understanding the Need for a Framework and Theory

- Jeff Hawkins, about artificial intelligence:
  - "We have an intuitive, strongly held, but incorrect assumption that has prevented us from seeing the answer. There is something that we believe that is obvious, but it is wrong."
- What is conspicuously lacking is a broad ***framework*** let alone a ***theory*** on how to think about digital financial reports.
- A **kluge** — a term from the engineering and computer science world — refers to something that is convoluted and messy but gets the job done.

Jeff Hawkins, Ted Talk, *How brain science will change computing*, Retrieved February 24, 2016; https://www.youtube.com/watch?v=G6CVj5IQkzk

Jeff Hawkins Ted Talk Statements

- "We don't need more data, we need a good theory."
- "Things look complicated until you understand them."
- "We have an intuitive, strongly held, but incorrect assumption that has prevented us from seeing the answer. There is something that we believe that is obvious, but it is wrong."
- "Intelligence is defined by prediction."
- "We experience the world in a sequence of patterns."
- "Real intelligence is a sequence of patterns."
- "Successful prediction is understanding."
- "Intelligence is making predictions about novel events."
- "Memory of high-dimensional patterns"
- "Memories are stored and recalled as a sequence of patterns"
- "Must be testable"
- "Must be buildable, if you cannot build it then you don't understand it."

Jeff Hawkins, Ted Talk, *How brain science will change computing*, Retrieved February 24, 2016;
https://www.youtube.com/watch?v=G6CVj5IQkzk

Financial Report Semantics and Dynamics **Theory**

- The purpose of this document is to propose and then prove that a defined set of semantics and dynamics or mechanics exists for a financial report, verify the correctness of these semantics and dynamics against publically available XBRL-based public company financial reports submitted to the SEC, and then to explain these semantics and dynamics in clear, logically coherent, consistent, and unambiguous terms in the form of a theory.

Charles Hoffman and Rene van Egmond, *Financial Report Semantics and Dynamics Theory*, Retrieved February 24, 2016; http://xbrlsite.azurewebsites.net/2016/Library/Theory-2016-02-22.pdf

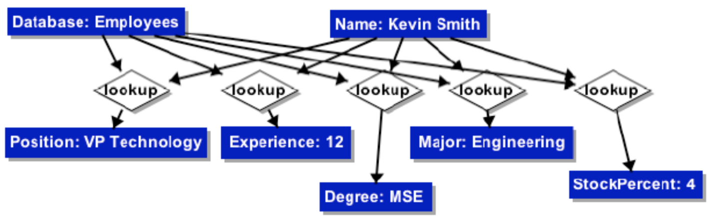
Financial Report Ontology

- An ontology is a representation of something that exists in reality.
- The financial report ontology is a formal ontology of a financial report.
- An ontology is a "window on reality". This financial report ontology is but one representative view of the world in machine-readable form.
- An ontology is a representation of universals. A universal is a generalization of instances. Instances exist only once.
- Machines such as computers need to be lead by the hand, they are like babies. The financial report ontology is intended to be able to guide computers, or more accurately computer software.
- The financial report ontology will help guide accountants creating a financial report or financial analysts working with one financial report or many financial reports.

Charles Hoffman, *Financial Report Ontology*, Retrieved February 24, 2016; http://www.xbrlsite.com/2015/fro/us-gaap/xbrl/Ontology/Overview.html

Database Values as Data

Name	Position	Yrs Experience	Degree	Major	Percent Stock
Karen Jones	VP Marketing	18	MBA	Marketing	3
Kevin Smith	VP Technology	12	MSE	Engineering	4
Keith Williams	VP Finance	15	BS	Accounting	3



Database Values as Information (data in context)

Name	Position	Yrs Experience	Degree	Major	Percent Stock
Karen Jones	VP Marketing	18	MBA	Marketing	3
Kevin Smith	VP Technology	12	MSE	Engineering	4
Keith Williams	VP Finance	15	BS	Accounting	3