Seamless Knowledge Management – A Proposal

Jian-Nan Zhang
University of Minnesota
May, 2013

0. Introduction – Story

Jean wants to learn a particular concept, he searches the concept on Google. Now the search engine presents hundreds of web pages – related lecture slides, forums, the Wikipedia article, and so on. After a glance, he decides to open the Wikipedia page. However, what Wikipedia gives him is a formal article, with numerous professional terms, jargons and complicated results, which is far beyond his understanding level of it. So he turns back to the search results, there are forums where a lot of people discuss similar topics, but a lot of them have conflict and there is no guarantee for the correctness of information. Now Jean is tired of doing search, because it is hard for him to get the **trustworthy information on his level**. The similar situation occurs not only on him, but also on other people. For a kid, it is especially hard to harness the information on the web, because little amount of pages are created for educating kids; for a professional, there are too many useless discussions and articles, some of them are wrong, others are meaningless for him to review. On the other hand, when the professional wants to study other topics, he encounters the same situation as a kid.

1. Introduction

Information now has gaps. Although the billions of pages on the World Wide Web have covered almost all aspects of human life and knowledge, the integration of all these information into knowledge is still a problem, and there are still dimensions of information hiding in human brains. There are two most important aspects of this problem, (1) The lack of information at different levels and types integrated together (2) Credibility of information is always doubtable. Even though enormous data models have been designed and ontologies have been applied to different knowledge content, the process of obtaining information, representing information, and processing information into knowledge at different levels is still not as good as it is expected.

On the other hand, the acquisition of knowledge is an accumulative process. When Jean wants to learn a particular piece of knowledge, he first needs to gain all types of intuitive observations, from the most basic level. For example, when he wants to know what a table is, he should first know something like “a table is made of wood”, “a table is usually used for dinner”, and look at some pictures about a table, etc. This type of knowledge can be concluded as commonsense knowledge. After that, he has the basics for more complex information. Then maybe he opens Wikipedia, and looks at some more formal content, such as the history, the types (bedside, refectory, drawing), and so on. Then if he is more interested in one particular type, he can find more information about the material, price, or if he likes design, he can get other articles about the different designs of table. But there are three key things in this process: first, Jean learns what he wants from the very easy and

intuitive to the very professional; second, the size of information is expanding exponentially from very general to very specific; lastly, the process of acquiring some knowledge depends on understanding other knowledge.

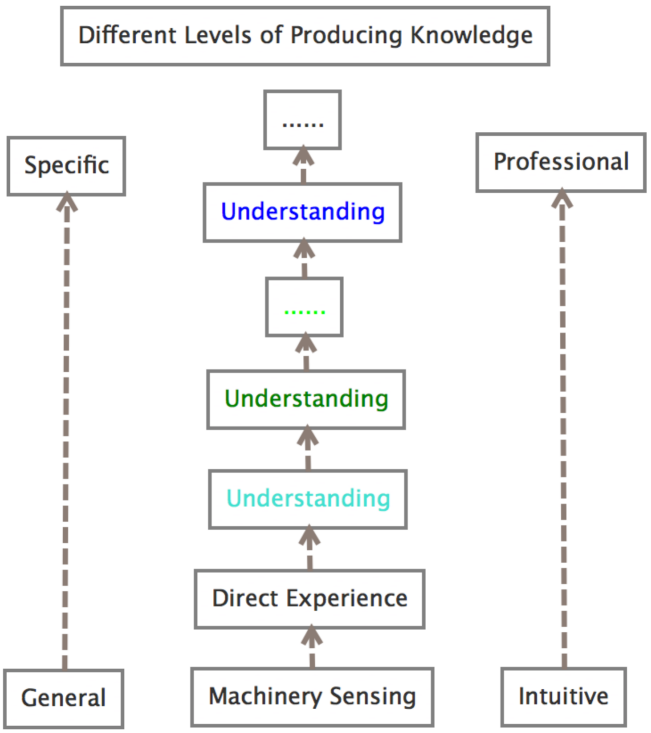


Figure 1 Knowledge Acquisition Process

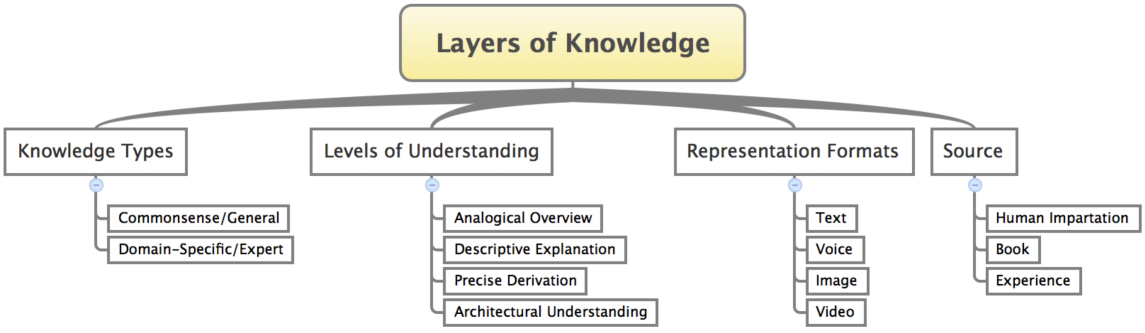


Figure 2 Different Perspectives of Knowledge Layers

The great diversity of the information determines no single ontology could be universal to represent all knowledge, and no single computational tools can be used to process all kinds of information, but a combination of all resources of all layers is the best way to utilize the current informational resource.

This paper proposes *Seamless Knowledge Management*, with an architecture for information organization, and a process to perform computation on different types of information.

In a broader sense, there are several dimensions to classify the content of knowledge.

1. Source of knowledge	Data Source
2. Type of knowledge	Ontology
3. Delivery of knowledge	Data Format
4. Processing of knowledge	Processing Runtimes
5. New knowledge	Applications/Outcomes

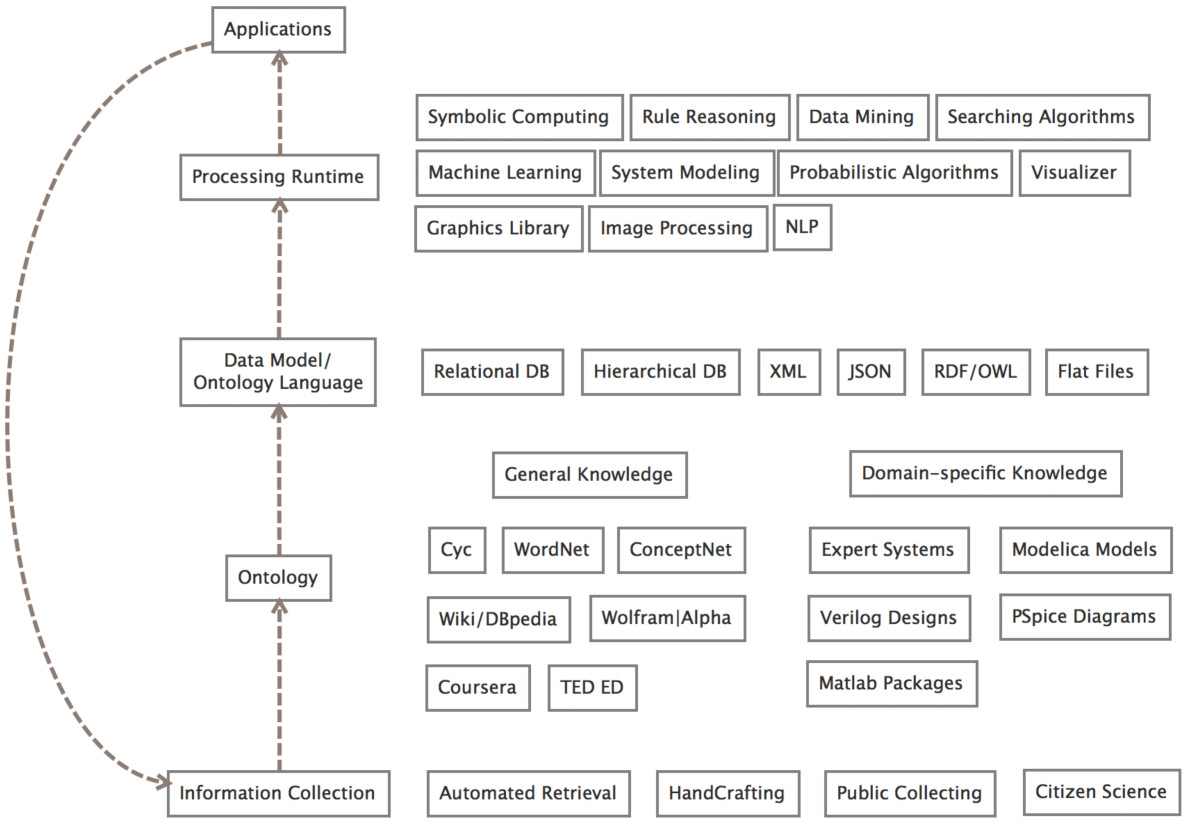


Figure 3 Seamless Knowledge Management -- Architecture

2. Source of Knowledge

Source of knowledge is crucial to its credibility and usability. However, both of these are very difficult to control: anyone can publish any kind of information on the Web, and there could be no absolute standard for the format.

2.1 Low-level Information Collection

2.1.1 Machinery Sensing

Sensors, cameras, scientific instruments, and computers generate huge amount of data every second, this type of data is “Raw data”, needs further processing.

2.1.2 Handcrafting

Knowledge Base like Cyc employs knowledge engineers to handcraft knowledge into a formal logical structure, and many other expert systems use this method to construct the knowledge base for further reasoning.

2.1.3 Plain Information

Plain information is the current information organization on the Web, with various blogs, discussion forums, etc. The content is hard to collect and process.

2.1.4 Citizen Science

The other way is to collect information from ordinary people. ConceptNet's commonsense data is collected from public, there are also other projects, such as the Geotag Libya, a post-conflict damage assessment platform. Citizen Science creates data to specific domains but with high quality from dedicated contributors. In a broader sense, Wikipedia is also included.

2.2 Middle-level Information Collection – Semantic Webs (Structured Information)

DBpedia
Wikidata
Wolfram|Alpha

2.3 High-level Information Collection
2.3.1 Education/Learning Activities
2.3.2 Professional Work
2.3.3. Academics

3. Combining Ontologies
There are no strict classifications of knowledge, but according to the extent of specialization, two types of knowledge can be distinguished, and cover known existing knowledge.

1. General Knowledge
   The scope of general knowledge is broad, an initial classification is:
   - Commonsense knowledge
     ConceptNet
   - General professional/academic knowledge
     Wikipedia, Cousera, Ted Ed
   - Experience knowledge
     Toyhouse blogs, stackoverflow

2. Domain-specific Knowledge
   Domain-specific knowledge contains various types of information. The specific expert systems, professional design works, source code, etc. All these information is not easy to be automatically retrieved because they go across too many types of data format, and a lot of them are even not open.

4. Delivery of Knowledge

5. Processing Runtimes

6. Knowledge Management System

A knowledge management system includes all five layers. The data flows between each layer in different formats. Clear workflow should be defined.

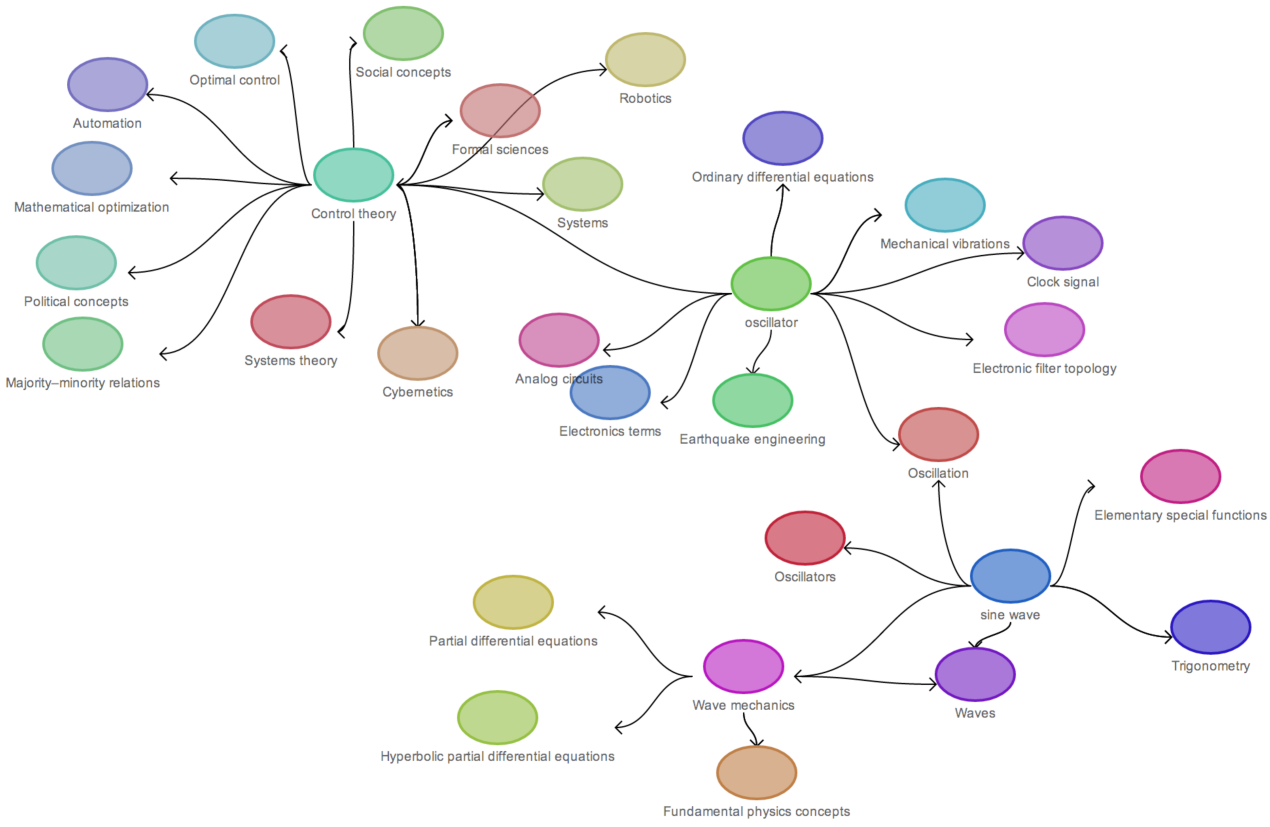
At the bottom layer, there is vast data but hard to process, to the top, knowledge should be highly structured and easy to process for different usages.

7. An Application – KnowledgeNet

The apps is purely using Javascript, which collects information from ConceptNet (Commonsense knowledge base), DBpedia (Wikipedia data), and Wolfram|Alpha (Computational knowledge engine).

It accepts any input keywords, and searches across three databases represent three types of data sources: commonsense knowledge, general knowledge, and computational knowledge.

For example, the following graph shows the result from keyword sequence "Wave → Wave Mechanics → Oscillator → Control Theory".



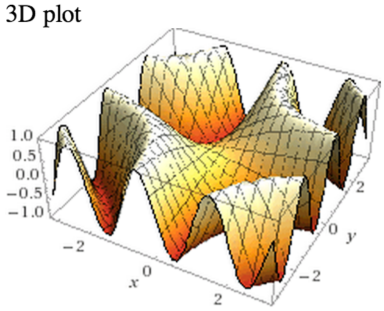
And there are short descriptions and links to Wikipedia for each of the search.

Wave mechanics
In physics, a wave is a disturbance or oscillation that travels through spacetime, accompanied by a transfer of energy. Wave motion transfers energy from one point to another, often with no permanent displacement of the particles of the medium—that is, with little or no associated mass transport. They consist, instead, of oscillations or vibrations around almost fixed locations. Waves are described by a wave equation which sets out how the disturbance proceeds over time. The wave equation is an important second-order linear partial differential equation for the description of waves – as they occur in physics – such as sound waves, light waves and water waves. It arises in fields like acoustics, electromagnetics, and fluid dynamics. Historically, the problem of a vibrating string such as that of a musical instrument was studied by Jean le Rond d'Alembert, Leonhard Euler, Daniel Bernoulli, and Joseph-Louis Lagrange. lookup wikipedia...

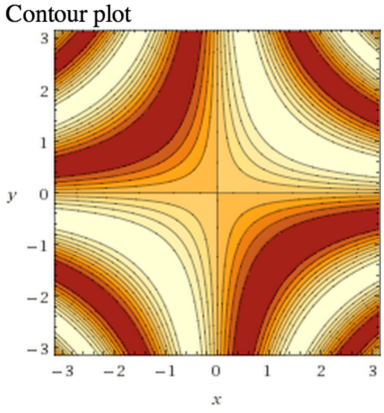
Wave
New Wave is a subgenre of rock music that emerged in the mid to late 1970s alongside punk rock. The term at first generally was synonymous with punk rock before being considered a genre in its own right that incorporated aspects of electronic and experimental music, mod subculture, disco and 1960s pop music. Frequency is the number of occurrences of a repeating event per unit time. It is also referred to as temporal frequency. The period is the duration of one cycle in a repeating event, so the period is the reciprocal of the frequency. For example, if a newborn baby's heart beats at a frequency of 120 times a minute, its period (the interval between beats) is half a second. Visible light (commonly referred to simply as light) is electromagnetic radiation that is visible to the human eye, and is responsible for the sense of sight. Visible light has a wavelength in the range of about 380 nanometres to about 740 nm – between the invisible infrared, with longer wavelengths and the invisible ultraviolet, with shorter wavelengths. A tsunami (-mee) is a series of water waves caused by the displacement of a large volume of a body of water, typically an ocean or a large lake. Earthquakes, volcanic eruptions and other underwater explosions (including detonations of underwater nuclear devices), landslides, glacier calvings, meteorite impacts and other disturbances above or below water all have the potential to generate a tsunami. Tsunami waves do not resemble normal sea waves, because their wavelength is far longer. Very high frequency (VHF) is the radio frequency range from 30 MHz to 300 MHz. Frequencies immediately below VHF are denoted high frequency (HF), and the next higher frequencies are known as ultra high frequency (UHF). The frequency allocation is done by ITU. These names referring to high-end frequency usage originate from mid-20th century, when regular radio service used MF, Medium Frequencies, better known as "AM" in USA, below the HF. lookup wikipedia...

When querying wolfram|Alpha, it gives more mathematical-based information about related topics:

3D plot



Contour plot



Alternate form

$$\frac{1}{2} i e^{-i x y} - \frac{1}{2} i e^{i x y}$$

Roots

$$x \neq 0, \quad y = \frac{\pi n}{x}, \quad n \in \mathbb{Z}$$

Properties as a function

$$x = 0$$

When more data comes, there will be more accurate and related results of different types

and different levels. Then the information acquisition can be more seamless, which makes everyone easier to retrieve.

There are several things to do:
1. Combine more databases to make information complete
2. Build more connections between different sources of data
4. Combine more processing runtimes (NLP, Learning Algorithms, etc.)

References
[1] Tim Berners-Lee, James Hendler, Ora Lassila. The Semantic Web. Scientific American
[2] Marvin Minsky. The Society of Mind
[3] Marvin Minsky. A Framework to Represent Knowledge
[4] Hugo Liu and Push Singh. ConceptNet – A Practical Commonsense Reasoning Tool-kit
[5] Francois Grey. Citizen Cyberscience for Africa