12-2015

Social Learning Systems: The Design of Evolutionary, Highly Scalable, Socially Curated Knowledge Systems

Nolan Hemmatazad
University of Nebraska at Omaha

Social Learning Systems: The Design of Evolutionary,

Highly Scalable, Socially Curated Knowledge Systems

By

Nolan Hemmatazad

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Information Technology

Under the Supervision of Dr. Qiuming Zhu

Omaha, Nebraska

December, 2015

Supervisory Committee:

Dr. Robin Gandhi

Dr. William Mahoney

Dr. Elliott Ostler

Social Learning Systems: The Design of Evolutionary,

Highly Scalable, Socially Curated Knowledge Systems

Nolan Hemmatazad, Ph.D.

University of Nebraska, 2015

Advisor: Qiuming Zhu

In recent times, great strides have been made towards the advancement of automated reasoning and knowledge management applications, along with their associated methodologies. The introduction of the World Wide Web peaked academicians' interest in harnessing the power of linked, online documents for the purpose of developing machine learning corpora, providing dynamical knowledge bases for question answering systems, fueling automated entity extraction applications, and performing graph analytic evaluations, such as uncovering the inherent structural semantics of linked pages. Even more recently, substantial attention in the wider computer science and information systems disciplines has been focused on the evolving study of social computing phenomena, primarily those associated with the use, development, and analysis of online social networks (OSN's).

This work followed an independent effort to develop an evolutionary knowledge management system, and outlines a model for integrating the wisdom of the crowd into

the process of collecting, analyzing, and curating data for dynamical knowledge systems. Throughout, we examine how relational data modeling, automated reasoning, crowdsourcing, and social curation techniques have been exploited to extend the utility of web-based, transactional knowledge management systems, creating a new breed of knowledge-based system in the process: the *Social Learning System* (SLS).

The key questions this work has explored by way of elucidating the SLS model include considerations for 1) how it is possible to unify Web and OSN mining techniques to conform to a versatile, structured, and computationally-efficient ontological framework, and 2) how large-scale knowledge projects may incorporate tiered collaborative editing systems in an effort to elicit knowledge contributions and curation activities from a diverse, participatory audience.

Dedication

This work is dedicated to my wife, Krystal, and daughter, Madelyn, without whom there are no stars in my night sky.

Contents

List of Multimedia Objects

1 Introduction

1.1 A New Breed of Knowledge Systems

Modern times have brought about myriad developments in the fields of automated reasoning and knowledge management (KM). With the advent of the World Wide Web came an explosion of interest in harnessing the power of linked, online documents for the purpose of developing machine learning corpora, providing dynamical knowledge bases for question answering systems, fueling automated entity extraction, and performing graph analytic evaluations, such as uncovering the inherent structural semantics of linked documents (a concept relied upon extensively in the development of search engines and document retrieval systems).

More recently, considerable attention in the wider computer science and information systems disciplines has been focused on the evolving study of social computing phenomena, predominantly those associated with the use, development, and analysis of large-scale online social networks (OSN's). Innovations such as these have generated a wealth of new content, at scales never before witnessed in a manner so readily accessible. Where the development of the internet and the Word Wide Web heralded an opening of the floodgates for traditionally more formal or articulated media contents, the emergence of OSN's and other social computing systems let loose a deluge of largely personal and sociocultural data, while in the process, cementing the internet's role as an exchange mechanism for all manner of ordinary human and, at times – owed to the recent *Internet of Things* movement – non-human discourse.

Forbus (2012) observed that sociality “might accelerate the bootstrapping of intelligent systems, and it could make them more effective collaborators. Hence it seems very important to explore”. This work looks at how nascent social computing concepts such as crowdsourcing and social curation, coupled with a wealth of new data made available by OSN's, can be harnessed in conjunction with traditional web mining techniques and information retrieval from structured data sources to create advanced, dynamical knowledge management systems, or *Social Learning Systems* (SLS’s). In bringing these elements together, we find that we can create generalized knowledge structures that are capable of providing access to a near infinite volume of diverse human knowledge.

In the SLS context, knowledge is not viewed as existing in disparate, static silos, but is instead seen as a perpetually evolving whole, where constituent entities may possess numerous fine relationships amongst one another. Where many existing knowledge bases compartmentalize information into separate data stores – at times bridging these divides superficially to present their contents in a unified manner (as is often the case with popular virtual assistant-style applications) – SLS’s invoke a generalized knowledge representation to unite concepts at the data storage level. SLS’s then adopt innovative social curation and collaborative editing techniques, in conjunction with rudimentary automated reasoning faculties, to couple data sourced from local, static data stores, third-party encyclopedic knowledge API’s, and Web- and OSN-mined contents, in a manner that is conducive to transactional knowledge reasoning tasks at scale, and which seeks to ensure information integrity.

1.2 From Knowledge to Knowledge Management

Cooley (1926) described knowledge as “a phase of higher organic evolution”, which seems to have emerged due to its purpose of “giving us adjustment to, and power over, the conditions under which we live". The act of *knowing*, as we're already well aware, serves to provide one with certain assurances regarding the nature of reality, perceived or otherwise. These assurances equip us with a non-arbitrary means of responding to our environment, thus affording a certain power, as Cooley suggests, to cope with various situations and conditions that may arise.

Though knowledge and the transfer thereof is a quintessential facet of human existence, the evolution of knowledge management and automated reasoning technologies has been a slow and trying process. Acknowledging such is not to promote a pessimistic outlook towards these fields, but rather to pay homage to the inherent challenges that must be faced throughout their advancement. Thagard (2009) noted that "whenever science operates at the edge of what is known, it runs into general issues about the nature of knowledge and reality". The cognitive and information theoretic sciences have, almost without exception, operated at the "edge of what is known" throughout the entirety of their existence, redefining many of the ways we view the world, and, more particularly, the mind. To further these domains and others that deal with similar subject matter is to face many underlying philosophical questions – both *ontological* and *epistemological* – pertaining to the nature of reality, to at times stretch the very boundaries of computability and technological capacity, and to grapple with a great number of practical questions relating to the implementation of new ideas; that is, the transformation from hypothetical to practical.

If we are to lay any meaningful foundations for next generation knowledge systems, the membership of an exhaustive problem space must be addressed in unison, and in a way that is highly pragmatic in nature. Failing this, we may be left with solutions that are fragmentary, or which insist on a divide between the realms of academia and practice (something we have been dutiful to avoid in the present work).

In a more focused view of knowledge management, Kankanhalli et al. (2003) identified two general varieties of KM supported by information technology: codification, in which "more explicit and structured knowledge is codified and stored in knowledge bases", and personalization, in which "more tacit and unstructured knowledge is shared largely through direct personal communication". In this work, we see how both varieties have been incorporated into the development of Social Learning Systems.

For these purposes, an SLS constitutes a system which, given a relatively limited initial knowledge base and primitive mechanisms for establishing relationships amongst concepts, can harness the contents of online social interactions and traditional information sources to form a more expansive and structured view of reality. SLS's then rely on direct interactions with their users to help vet their information contents and inherent taxonomical structure through a tiered system of collaborative editing, involving both social curation activities and traditional KM editing functions.

The end goal, therefore, is to model and deploy generalized knowledge systems that are capable of transactional reasoning at scale, the likes of which are required for evolutionary advancements along several prominent veins of artificial intelligence and knowledge management research. As just one example, in our work, we've reviewed the well-established field of open domain question answering (QA), due to its ability to benefit directly from advances in general purpose KM technologies and social curation techniques, in addition to its inclusion within the SLS model as a means for directing Web and social data mining tasks.

The exploration of these ideas is met with several fundamental philosophical challenges, and novel solutions to a variety of practical problems proved necessary to engage these obstacles in a meaningful way. More detailed coverage of these challenges and the theoretical underpinnings relied upon to help address questions of a philosophical nature are elaborated upon in the latter sections of this work.

1.3 Objectives Statement

This objectives statement, in much the same manner as a corporate mission statement, is not intended to be exhaustive, but rather to express, succinctly, the nature of these research endeavors. Thus, our objective was to detail a model for a new kind of knowledge management system that utilizes data from the World Wide Web, online social networks, and existing structured data sources (both local and remote), in conjunction with basic automated reasoning, social curation, and collaborative editing mechanisms, to provide an evolutionary framework for conducting transactional knowledge tasks at scale, and in a computationally-efficient manner.

1.4 Key Questions

There are several questions that this research seeks to address, in some cases directly and in others, indirectly. While many are novel in the more general sense, each has facets that are unique within the context of our work. With that said, the core question that this work seeks to address – and which has been broached previously in this text – is as follows:

1. How can we integrate web- and OSN-mined data, in conjunction with local, structured data sources and data retrieved from third-party knowledge API's within a generally-applicable, computationally-efficient ontological framework?

Rao (2003) observed that “a general-purpose taxonomy would probably be less useful than appropriately specialized or even private taxonomies. Focused taxonomies are likely to make finer-grain discriminations within topics in more specialized collections, and are also likely to better match the language and the purposes of specialized users and uses”. The inherently “fuzzy” nature of social interactions and Web-mined data mandates that this framework be versatile in accommodating information in many forms, from a variety of sources, and without any preference to a particular knowledge domain. Information from more informal or unstructured sources is often absent from existing knowledge bases, much less information passively gleaned from ordinary discourse (although accumulating information in this format may be common amongst human actors). This phenomenon is observable in the domain of QA, where according to Soricut and Brill (2006), “with very few exceptions, most of the work done… focuses on answering factoid questions”, while “the world beyond… is largely unexplored”. This, in essence, creates a gap between the kinds of knowledge that are the subject matter of KM systems, and the larger realm of knowledge that people concern themselves with in their day-to-day lives, which often incorporate information exchanged through personal interactions and other unstructured mediums.

A number of additional themes that this work acknowledges in a secondary, less exhaustive capacity, are as follows:

2. How can we integrate primitive automated reasoning, digital curation, and collaborative editing faculties within this new class of systems in order to help vet the integrity of their resulting knowledge bases, and supplement their information

contents? Digital curation as a whole is still a widely unexplored research area, and little work has been done to integrate such faculties at both the structural and intra-entity levels within dynamical knowledge systems. With that said, this work should serve as a novel demonstration of the integration of these components, a task that is particularly interesting within the context of a large-scale knowledge system where data may be sourced from the World Wide Web and social contexts. As Guarino (1995) notes, “a knowledge base will acquire a value *per se*, only to the extent that the knowledge it contains is in fact true, such as to correspond to the world beyond the knowledge-base”. In general or open domain knowledge bases, the world outside the knowledge base is, in fact, the *world at large*, and with a plasticity of means, the integrity of system outputs must rightly come under scrutiny. More to the point, a system that incorporates information from less authoritative sources must also incorporate greater means of vetting the truthfulness of data sourced through these less definitive means.

3. Lastly, how might QA functionality be integrated within Social Learning Systems in order to direct Web and social knowledge retrieval efforts in a manner that is conducive to achieving the overall objectives outlined previously? Kwok et al. (2001) cited several obstacles to performing question answering over the web, including “forming the right queries”, noise, dealing with falsehoods or false positives, and resource limitations. For complex questions, Soricut and Brill (2006) have additionally speculated that “it is extremely unlikely that any type of question reformulation will increase the chance of finding the document containing the answer”. In moving forward from such historical hurdles, how can

we simultaneously expand the horizon of data sources we interact with, while solving the problem of honing in on the correct information existing within particular knowledge sources (or even establishing that the correct information can be located to begin with)?

The rest of this paper is organized as follows: first, we describe the conceptual framework within which we construct an operational view of the issues under consideration in this work. Following this, the methodology of our work is reviewed, including its associated constraints and limitations. Next, we take an in-depth look at the anatomy of a Social Learning System, before seeking to establish the effectiveness of Social Learning Systems with regards to their computational efficiency and suitability for Web-scale transactional knowledge tasks. Finally, we offer some suggestions for future research directions for those who may be interested in conducting related studies.

1.5 Legal Notice

This document includes several screenshots of a demonstration SLS developed externally to the author's university affiliation. All such items are under copyright by their respective rightsholder(s). These screenshots may depict text or media owned by various individuals or organizations, which are presented here under the Fair Use doctrine. This work serves only as an explication of a computational social learning model, while also providing general guidelines for the development of future systems of a similar nature. As such, all intellectual property associated with any specific Social Learning System described within this document remains the exclusive property of the associated rightsholder(s).

2 Conceptual Framework

The current section serves to outline a conceptual framework that can help buoy a cogent way of thinking about SLS's. We begin by relating a historic metaphilosophy of the mind[1], incorporating the views of seminal thinkers on the nature of the mind and the knowledge acquisition process. Following this, we establish a theory base for learning, extending from the roots of empiricism, which incorporates sociality as a primary motivator for the procurement of new knowledge. We then provide definitions for key terms used throughout this text, as well as example usage scenarios for a practical Social Learning System.

[1] As this work deals predominantly with the topic of social learning, notably absent from its main contents is a treatment of the histories of relevant computer science cognition disciplines. We would be remiss to avoid incorporating such a discussion altogether, however, and so readers are encouraged to look to *Appendix A* for a more thorough treatment of these fields, as well as *Appendix B* for pointers to other related disciplines.

2.1 Historic Metaphilosophy of the Mind

Principally by way of his 17th century collection of writings entitled, "An Essay Concerning Human Understanding", the work of John Locke has proven critical to many historically-defined views of the mind. In these texts, Locke set the stage for the contemporary – and highly controversial – notion of the blank slate, or *tabula rasa*, mind. Locke's views of the mind as being innately without content have been challenged for centuries by myriad sources offering their own alternative views of the mind, often incorporating evolutionary mechanics or Piaget's "genetic epistemology" as the basis of their position.

However, their arguments, as a whole, rest a great deal on subtle misinterpretations of Locke's writings, often presupposing that the "content-free" mind must somehow also be one void of innate functionality. In fact, what Locke had written in the original essay was that the individual "by the use of their natural faculties" – which we must not assume would constrain the influence of genetics or the evolution of neural physiology, as little known at the time as they were – can "attain to all the knowledge they have, without the help of any innate *impressions* [emphasis added]" (Book 1, Chapter 2). This, by reasonable interpretation, is to suggest that one does not need to understand that something may appear as *X* or occur in *Y* manner in order for them to possess the necessary mental faculties for registering how something has appeared or occurred, as well as to be able to refer back to that particular occurrence at a later date.

Locke observed, too, that "in ideas… the mind discovers that some agree and others differ, probably as soon as it has any use of memory; as soon as it is able to retain and perceive distinct ideas". This statement, taken alone, helps to elucidate Locke's meaning of the content-free mind as one of functional *potential*, capable of processing and assimilating an endless range of external stimuli. This process of perception and the contrasting of distinct ideas, Locke observes in Book 4, is the foundation of all knowledge. Viewed in this light, it becomes more difficult to see how Locke's notions of the mind and more recent philosophies or empirical developments in the cognitive and neural sciences should have ever been at odds with one another.

In some ways, this view of the mind can be traced back even earlier, to the 13th century and the related philosophy of Saint Thomas Aquinas. Most important to our own pursuits is Aquinas's Commentary on Aristotle's *De Anima* ("On the Soul", originally composed in the 4th century, B.C.), and in particular, Book 3, Chapter 4, where the philosopher considers Aristotle's treatment of general intellectual functioning. There, Aquinas describes the intellect as a kind of sensory device, or a *potentiality*, capable of the reception of "intelligible objects". Setting aside potential differences in the perceived composition of the mind or its designation as a spiritual force, we can see that this view aligns well with the more recent Lockean tradition.

Both of these views are in harmony also in the sense that, when dealing with impressions upon the understanding (per the Lockean view) or perceptions of the intelligible by the intellect (per Aristotle/Aquinas), the focus appears to be on the content of our physical existence; "real" things, that is, or at least things conceived of in the mind through no form of divine intervention. What becomes understood or absorbed into the intellect through experience lays the foundations for future knowledge.

This is of crucial significance to the empiricist paradigm. Empiricism is, as Bechtel (2009) notes, "the idea that all knowledge is rooted in sensory experience". Though Aristotle may have been an early empiricist in his own right, it would seem that contemporary notions of empiricism date back somewhat later, to perhaps the 3rd century, B.C., and to the Empiric school of medicine. Among the sect's membership was at least one particularly outspoken protagonist of experiential methods of understanding, Sextus Empiricus. As Chisholm (1941) wrote, "although the true sceptic should question any proposition which refers beyond that which is immediately before him, it is impossible, according to Sextus, to be sceptical about the given itself". This mode of empiricism (and that associated with Aristotle) likely followed in the footsteps of more mechanistic views of reality, such as the atomic theory of Leucippus and Democritus.

Regardless of what chain of events manifested to give empiricism and scientific methods of understanding the foothold they quickly developed, it was a sequence that would continue unfolding even in contemporary times. John Dewey, whom among other things was a respected educational reformer, stressed that one should "discriminate between beliefs that rest upon tested evidence and those that do not", and to be "on guard as to the kind and degree of assent yielded" (Dewey, 1910). These views would become central to the American educational philosophy, embedding empiricism yet more deeply into the fabric of various intellectual pursuits.

2.2 Theory Base for Learning

> "Even the simplest perceptions of form or extent, much more the exact perceptions of science, far from being mere physical data, are the outcome of an extended process of education, interpretation, and social evolution" (Cooley, 1926).

For inspiration and philosophical grounding for the "learning" faculties of the class of systems described in this work, we look to social learning theory, an evolution of the empiricist philosophy of mind. Bandura (1971) wrote that, "in the social learning view, man is neither driven by inner forces nor buffeted helplessly by environmental influences. Rather, psychological functioning is best understood in terms of a continuous reciprocal interaction between behavior and its controlling conditions". In this system, learning is established in large part due to the observed modeling of appropriate behaviors and differential reinforcement for the exhibited behaviors. In the earlier work of Cooley (1926), this variety of knowledge was referred to as "social knowledge", in contrast to "material knowledge", which derives more directly from the empirical senses.

Cooley also paid homage to the introspective mind, through which “we come to know about other people and about ourselves by watching not only the interplay of action, but also that of thought and feeling”.

Coincidentally, the notion of differential reinforcement has already found a home in the field of automated reasoning, wherein Wolpert and Tumer (2008) suggested that “because [reinforcement learning] generally provides model-free and 'online' learning features, it is ideally suited for the distributed environment where a 'teacher' is not available and the agents need to learn successful strategies based on 'rewards' and 'penalties' they receive from the overall system at various intervals”. How this variety of reinforcement learning manifests within the specific context of Social Learning Systems is discussed later.

In his work, Wenger (2000) observed that “in a social learning system, competence is historically and socially defined… Knowing, therefore, is a matter of displaying competences defined in social communities”. Wenger notes, as well, that “socially defined competence is always in interplay with our experience”, and that, from this interplay, we realize the process of learning. For knowledge systems incorporating automated reasoning and social curation faculties, this view is highly appropriate, as it relates well to both the natural computational sequence of events that transpires in ordinary system activities, as well as the role that social interactions play in shaping a more highly refined understanding of the world.

Truly learning from an observed behavior, however, depends upon the successful assimilation and recall of acquired information. In social learning theory, this is known as a *retention process*, without which, one is unlikely to be influenced by modeled behaviors (Bandura, 1971). In an automated reasoning context, this might suggest that discovered information and computed assertions cannot be ephemeral in nature, but must instead be used as a foundation upon which future knowledge claims are developed or assessed.

The variety of social learning described here has witnessed adoption beyond the realms of education, psychology, and the computer sciences. In an economic context, Arrow contrasted "methodological individualism" with the inherently social market, noting that "individual behavior is always mediated by social relations", and that "these are as much a part of the description of reality as is individual behavior" (1994). Similar applications of social learning theory have appeared in works relating to criminology, health sciences, communication, business administration, and various other domains. The widespread employment of the social learning view in academic contexts helps affirm its suitability as a theoretical learning base in the development of social computational knowledge systems, such as the Social Learning Systems described here.

Importantly, intelligence, while a "social phenomenon" (Mataric, 1993) fostered in a context of shared norms and understanding, must still be made manifest by the individual. It is within the minds of individuals, after all, that the requisite ascension from information to knowledge occurs through interpretation, as well as the anchoring of information in "the beliefs and commitments" of the host individual (Nonaka et al., 2000). Reasoning systems, too, must possess internalized processes for information interpretation, and some structured set of beliefs or commitments (a framework in which to accord the information) if the end goal of social learning is to be meaningfully achieved.

2.3 Key Terms and Definitions

It is assumed that the readers of this work will have been previously acquainted with many of the terms used throughout its contents. For a few concepts, however, the meanings may not be as readily apparent, or may possess special characteristics worth highlighting with respect to the current research endeavors. These items are elaborated upon in the subsections that follow.

(Un)Structured Data. The dichotomy between structured and unstructured data is a well-known phenomenon in computer science and related fields. Structured data has been defined as "any set of data values conforming to a common *schema* or *type*" (Arasu and Garcia-Molina, 2003), such as a table containing values that follow implicit or explicit patterns, while unstructured data "consists of any data stored in an unstructured format at an atomic level. That is, in the unstructured content, there is no conceptual definition and no data type definition" (Weglarz, 2004). Extending from this, it should be apparent that the text you are reading at present constitutes predominantly unstructured data. A significant subclass of unstructured data has been referred to as *tacit knowledge*, which in KM contexts represents all variety of supposedly non-quantifiable knowledge, especially that of a social or sociocultural nature (Linde, 2001). In terms of practical complexity (with the exception, perhaps, of structured data at scale), unstructured data has received the lion's share of researcher attention, due to its inherently lower degree of amenability to traditional computational operations.

As Rao (2003) explains, however, "neither content nor knowledge work is truly unstructured... Content, despite often being called 'unstructured data,' is shaped – first, by intrinsic aspects of representation and expression and, second, by the social context in which it is produced and consumed". This bears some significance for computational efficiency considerations, where relying on any semblance of structure within the available data can greatly impact performance (Buneman et al., 1997). This concept has been relied upon extensively for research involving automated entity extraction and linguistic analyses.

Knowledge Representation and Ontologies. Knowledge representation, in its simplest sense, relates to the systematic process by which we schematize, store, and later apply formal rules upon information for use in larger knowledge management-oriented applications. As Davis et al. (1993) remind us, however, these representations are not data structures. Knowledge representations are implemented through data structures in much the same way that a piece of software is implemented in a programming language; the data structure or language provides more general constraints, but additional assumptions or rules may be adopted within the application or knowledge representation which render them inequivalent at a conceptual level.

Ontology, broadly-speaking, is "the study of the organization and the nature of the *world* independently of the form of our knowledge about it" (Guarino, 1995). In the information sciences, ontologies serve as a means of recording information about the world in a way that promotes reasoning, such as through employing data structures conducive to mapping relationships between concepts, as well as individual-level attributes or properties. They work, in this sense, as a "shared conceptualization of a domain" (Lee et al., 2011). In adopting these ontologies, one must make a set of *ontological commitments*, which are "in effect, a strong pair of glasses that determine what we can see, bringing some part of the world into sharp focus at the expense of blurring other parts” (Davis et al., 1993). Just as a human conceptualization of the world presupposes a certain set of assumptions, an ontological knowledge representation does, as well, if only due to technical constraints.

In their 2007 article, Brewster and O'Hara identified several contemporary challenges relating to ontologies, including language ambiguity and constraints on language expressiveness, the impossibility of achieving perfect fidelity in conceptual encoding, inherent and deep-rooted commitments to certain epistemological viewpoints inherent in schematization, a structural bias towards computability, contingencies upon human expression, trade-offs between expressiveness and usability, knowledge currency restrictions, maintenance requirements, lack of universal applicability, inflexibility or rigidity in knowledge encoding, and others. Several of these complexities may be attributed to the fact that, as Guarino (1995) observes, AI researchers – the founding fathers of information science ontology development – “seem to have been much more interested in the nature of reasoning rather than in the nature of the real world”. In other words, it can be said that ontologies have been historically more deeply seated in epistemology, as opposed to their namesake philosophical interest.

Social Computing and Social Networks. Social computing has been broadly defined as “the use of computational devices to facilitate or augment the social interactions of their users, or to evaluate those interactions in an effort to obtain new information” (Hemmatazad, 2014). More succinctly, social computing “extends the scope of usage of information and computing tools to the realm of social endeavors” (Parameswaran and Whinston, 2007a). One of the best known examples of social computing in the real world can be found in online social networks (OSN’s, or simply *social networks*). Boyd and Ellison (2007) defined these networks as “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system”. Social networks serve as an outstanding demonstration of social computing principles in the real world, and are among the most highly utilized computational systems in operation today.

Digital Curation. If one were to look for a traditional definition of a *curator*, they might find them referred to as “the stewards of our history… who typically manage and take care of artifactual collections at cultural heritage institutions and who organize exhibits in galleries” (Liu, 2010). In the digital sense, curation may refer to any of a variety of proactive measures taken to ensure the reliability and accessibility of information resources over time. The DigCCurr project, a collaborative research initiative to develop a digital curation curriculum for graduate students at the University of North Carolina at Chapel Hill, is just one illustration of the increasing need to forge a more rigorous understanding and practical appreciation of the field of digital curation. In the guiding principles for the DigCCurr project, it was noted by Lee et al. (2007) that “digital curation activities span the entire lifecycle of digital resources”, making it worthy of integration consideration for large-scale knowledge management projects early on.

Even more recently than the DigCCurr project, Yakel et al. (2011) discussed the development of a digital curation curriculum at the University of Michigan's School of Information, with a specialization in information preservation in an archival context. This underscores a growing need for developing best practices for managing knowledge repository contents in light of their more temporal side, or to acknowledge that information wields the potential to change or evolve over time. Adapting this to the context of Social Learning Systems suggests that we must find ways to ensure the utility of what is true today, without sacrificing the fidelity of what has been true in the past or what may be true tomorrow.

To ensure the accessibility and reliability of knowledge contents, Social Learning Systems employ a tiered system of collaborative editing, with a specialized form of digital curation being the lowest and most accessible level of participation. In the SLS context, digital curation is inspired by the notion of the "wisdom of the crowd". This inspiration is so considerable, in fact, as to merit a rebranding from *digital curation* to *social curation*. Social curation bears some resemblance to the more abstract and better known concept of crowdsourcing, a "model that harnesses the creative solutions of a distributed network of individuals through what amounts to an open call for proposals" (Brabham, 2008). Crowdsourcing, in turn, befits many of the characteristics of social computing at large, as its effectiveness stems in many ways from the increasing propensity of communities to drive innovation "from the bottom up", as well as to take "ownership of experience, economic value, and authority" in the place of more established institutions (Wang et al., 2007).

Notably, our treatment of social curation in this text will contrast with that of Duh et al. (2012), who refer to the term as the "process of remixing social media contents for the purpose of further consumption". In their work, the essential emphasis of social curation was on the *curation of social contents*, whereas here, emphasis is placed on the *social curation of contents*.

Within this understanding of social curation, a form of collective intelligence or "collective problem-solving ability" (Heylighen, 1999) emerges, where "content curation communities" arise in an effort to "aggregate, validate, and annotate" contents (Rotman et al., 2012), or in our case, the knowledge artifacts of a Social Learning System. To wit, social curation can be a very appealing technique to incorporate into a knowledge management toolkit. Within an SLS, social curation works hand-in-hand with more traditional forms of collaborative editing, the likes of which have already achieved a high degree of success in other large-scale web-based KM projects, such as the popular Wikipedia project from Wikimedia, or the collaboratively-created graph database, Freebase (Bollacker, 2008). Though recently rendered defunct by its maintainers, Freebase was a particularly relevant source of inspiration for the present work, due to its reliance upon the taxonomic classification of knowledge artifacts, a quality that has proven essential in the construction of an SLS.

Knowledge Management (KM) and Knowledge-Based Systems (KBS). KM's and KBS's are, in so few words, systems that manage or rely upon the existence of knowledge artifacts. The former – knowledge management systems – predominantly serve the purpose of facilitating the sharing of knowledge across a constituency of human actors (Evermann, 2005; Singh and Kant, 2008). The latter – knowledge-based systems – exploit existing knowledge stores for empowering automated reasoning faculties (Evermann, 2005). The knowledge management process, which is of somewhat greater import to the present work, incorporates four actions: knowledge gathering, organization and structuring, refinement, and distribution (Benjamins et al., 1998), each of which must be present within a functional Social Learning System.

The terms described in this section account for only a small subset of concepts involved in the development of Social Learning Systems, an explication of the full breadth of which remains outside the focus of this work. It is our hope, however, that this background will prove useful in the methodology discussions that follow.

2.4 Example Usage Scenarios

Having introduced a conceptual framework within which to reason about Social Learning Systems, as well as elaborated upon some of their central concepts and inspirations, in the space that follows, we examine a number of potential usage scenarios that SLS's might one day find themselves a party to, in order to focus a more practical lens on the systems being proposed. Many readers will have already been acquainted with the relevant application domains, while for others, the below exploration will help to demonstrate the versatility of the systems being discussed, while at the same time showcasing their value for future research and development efforts.

Computational knowledge and reference systems. Social Learning Systems may be used as computational knowledge engines or reference systems, in the vein of Wolfram|Alpha or the collaborative encyclopedia, Wikipedia. These systems can offer substantial value as computational knowledge platforms due to their lack of absolute reliance upon existing structured data sets, thereby allowing their information contents to be much more dynamic in nature, and thus, to some degree, escaping a traditional limiting factor that is present in many existing applications in this area. The inherent social facets of these systems also align well with the spirit of existing collaborative knowledge platforms, where reliance upon the wisdom of the crowd represents a large part of the underlying appeal of using such services.

Fact-checking and question answering. Social Learning Systems find referent value in the discipline of question answering, and as a result, bring the same utility and value proposition of question answering systems to web and OSN scales, making them highly desirable platforms for traditional QA and fact-checking tasks. SLS's further expand the state of QA systems via their integration of social curation, and through the employment of a unifying ontology that allows for much broader information integration, without an overwhelming burden being levied upon overall computational efficiency.

Reputation management and sentiment analysis. Admitting informal social discourse into the system's structured knowledge backbone provides the interesting side effect of Social Learning Systems being well-suited for the exploration of online reputation and sentiment analyses. In many ways, subjective data constitutes potentially useful information, as it can shed light on the interpretations and predispositions of a larger social group. Further, ingrained confidence faculties provide a convenient and practical mechanism for assessing the relative diversity of opinions that may exist surrounding particular subjective claims.

Intelligent or artificial personal assistants. Being jointly capable of reasoning from existing structured data sources as well as social discourse, Social Learning Systems are well-poised for the performance of automated personal assistant-style information tasks. This is important due to the fact that, as mobile technology capabilities continue to grow at a rapid rate, the use of these artificial assistant technologies – and future variants offering even more advanced and computationally-intensive functionality – may become more commonplace amongst technology consumers.

3 Methodology

3.1 Research Methodology for Conducting the Study

In conducting the current research work, we have adopted the design science research paradigm. As Cross (2001) explains, "design science refers to an explicitly organized, rational, and wholly systematic approach to design; not just the utilization of scientific knowledge of artifacts, but design in some sense as a scientific activity itself." Elaborating on this notion, Hevner et al. observed that design science "seeks to create innovations that define the ideas, practices, technical capabilities, and products through which the analysis, design, implementation, management, and use of information systems can be effectively and efficiently accomplished." This underlines a strong focus on utility and inventiveness in areas where "existing theory is often insufficient" (Hevner et al., 2004).

Vaishnavi and Kuechler suggest that research in design science is distinct from the act of design in isolation due the inherent focus on the "production of interesting (to a community) new knowledge". The authors further note that design science research is novel due to its "intellectual risk", or "the number of unknowns in the proposed design which when successfully surmounted provide the new information that makes the effort research and assures its value" (Vaishnavi and Kuechler, 2004).

In Peffers et al. (2007), an iterative Design Science Research Methodology (DSRM) process consisting of six dominant activities is presented. The stages of this process are summarized in the space below:

1. "*Problem identification and motivation.*" In this stage, the researcher defines a problem explicitly and provides a rationale for why the problem (and the corresponding quest for a solution) are important.
2. "*Define the objectives for a solution.*" Here, a researcher describes the specific goals of a potential solution to the problem, based on what is both "possible and feasible" at the time, and accounting for situational constraints.
3. "*Design and development.*" This stage involves the actual creation of the solution.
4. "*Demonstration.*" In this step of the process, it is shown how the developed solution addresses the defined problems through actual usage.
5. "*Evaluation.*" Empirical analyses can be performed at this stage in order to assess the solution's proficiency at addressing the problem, beyond simply showing how the solution works.
6. "*Communication.*" In the final stage, findings and an explication of the process for developing a solution are related to a larger knowledge community with interest in the related subject matter.

This process is presented visually in Figure 1.

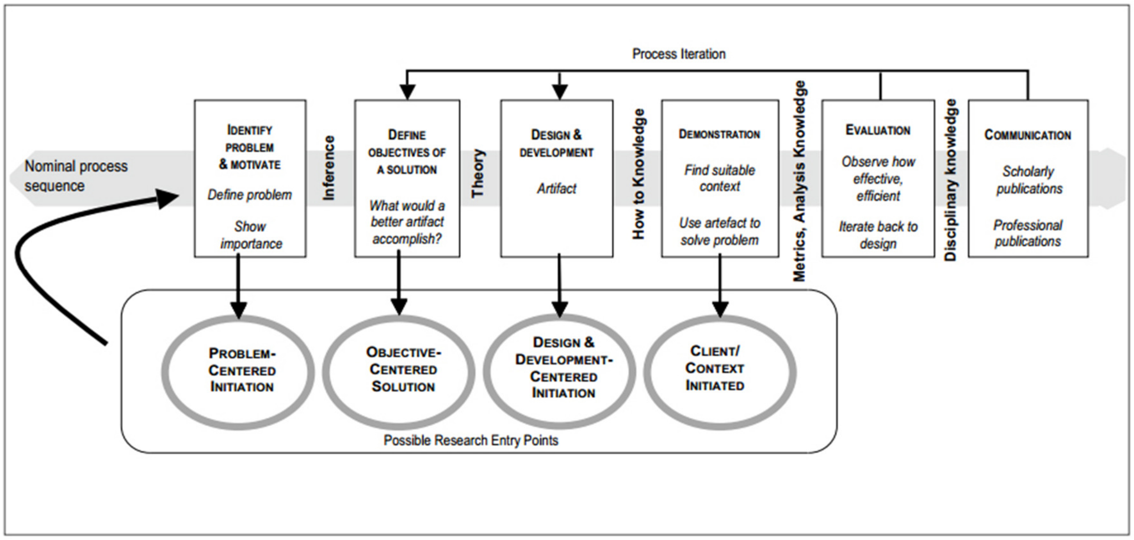


Figure 1. A DSRM Process Model (Source: Peffers et al., 2007).

Our research (and the associated application development work) adhered to the process identified by Peffers et al., with the addition of an intermediary stage falling between *defining objectives of a solution* and *design and development*: that is, the construction of an initial theory base upon which we ground the development of an appropriate solution, as has already been expounded upon earlier in the *Conceptual Framework* section of this paper. It is important to note, however, that in traditional design science research, reliance upon an existing theory base has been given relatively less significance in contrast to most popular research methodologies, due to the highly applied nature of the design science philosophy. For our work, the inclusion of a theory base was intended to help inform readers of the nature of social learning, as well as to provide some initial background on relevant philosophies of the mind, so that they may better understand some of the design choices that have been made throughout system development.

3.2 System Development

Several development paths intersect in the construction of a practical Social Learning System. Perhaps the most noteworthy among these are the development of subsystems specifically designed to handle structured data retrieval from primary information sources, question answering operations, automated reasoning and consensus tasks, data modeling and storage, digital curation, and information presentation. Several of these subcomponents are detailed more thoroughly in the sections that follow.

Question Answering Functionality. Waltz (1978), in expressing the development goals of a question answering system in the field of aviation, recognized the following major requirements for QA systems:

1. the ability to accept and work with natural language inputs,
2. explicit answer generation (i.e., in contrast to a list of potential sources of knowledge),
3. concessions for tolerating minor errors,
4. the use of "clarifying dialogues" for purposes such as resolving ambiguities,
5. ease-of-use and user experience accommodations, and
6. extensibility to add new functionality, or to expand on existing knowledge sources.

In addition to the work of Waltz, Kwok et al. (2001) identified three high-level technical components of a QA system: 1) an information retrieval engine (for their purposes, a traditional search engine), 2) a query formulation mechanism, and 3) an answer extractor. While a number of the requirements for a QA system outlined by Waltz may be unnecessary for the development of an SLS (due to its purpose not being restricted solely to QA-related tasks), each of the components identified by Kwok et al. prove essential for the development of a QA system exhibiting even modest complexity. As such, a brief description of each is provided below:

- The **information retrieval engine** consists of one or more knowledge bases featuring a public interface or gateway for accessing data in a computationally viable manner. For the SLS discussed in the current work, sources consisted of structured local knowledge bases, web-based search engines, external knowledge repositories, and online social networks. For QA tasks in particular, only information retrieved from web-based and social sources was considered, while working with data from other information sources was left to additional information retrieval and parsing faculties of the SLS.
- A **query formulation mechanism** accepts user input, processes the input, and transforms it into appropriate queries for passing along to an information retrieval engine. Kwok et al. (2001) identify multiple means of query transformation that go beyond merely adapting a request to a specific query language, including: verb conversion (e.g., from *did visit* to *visited*), query expansion (e.g., finding attributive nouns of adjectives), noun phrase formation (i.e., maintaining structure of compound nouns), and transformation, or syntactically rearranging the elements of a question into "equivalent assertions" (this last method being that which is adopted within the demonstration application presented here). Another tool often employed in query formulation is *word sense disambiguation*, which Sebastiani (2002) defines as "the activity of finding, given the occurrence in a text of an ambiguous (i.e., polysemous or homonymous) word, the sense of this

particular word occurrence". While query transformation can be highly complicated, requiring a time-intensive development process, it's been demonstrated that the complexity of question rewriting tasks can be greatly reduced in environments with a large number of information sources, due to the increased likelihood of answer matches being expressed in more diverse ways (Dumais et al., 2002). This is significant due to the fact that SLS's, being able to work with information from a wide variety of sources, boast one of the largest assemblages of source material of any contemporary knowledge systems.

- In the system developed by Waltz (1978), user request processing is comprised of four stages: the parsing and query generation stages, which fall under the earlier heading of *query formulation mechanisms*, as well as evaluation and response, which constitute principle constituents of the **answer extraction** component, where information is extracted from output generated by the retrieval engine and eventually displayed, in part or in whole, to the end user. In the demonstrated SLS, answer extraction was handled primarily via basic pattern matching tasks and linguistic template conformance tests, with subsequent linguistic clustering to group together related assertions.

Additionally, though not incorporated in the work of Kwok et al., a tiered social curation and collaborative editing architecture constitutes a significant technical component of the QA faculties implemented within the Social Learning System artifact presented in this work. In the social curation component, which initially jolts to life following the answer extraction and presentation stages of the QA process, users are able to interact with the system in order to help certify existing knowledge claims, including factoids pertaining to a specific knowledge entity or relationships from one entity to another.

Automated Reasoning and Consensus Tasks. Certain automated reasoning faculties prove necessary in order to establish an alignment of knowledge claims across multiple information sources. This can be thought of as the "consensus" component of a Social Learning System. In related works, voting schemes for establishing information consensus have been adopted and discussed in great depth throughout the literature on collective intelligence (see, e.g., Malone et al., 2009). Schemes such as these reside within a larger class of *consensus tasks*, which have been defined as a means of identifying "a hidden state of the world by collecting multiple assessments from human workers", with the additional quality that the "state of a consensus task at any time step is defined by the history of observations collected for the task" (Kamar et al., 2012). In our context, as in the case of Malone et al., the notable distinction is that these consensus tasks need not be conducted entirely by humans, but can be conducted autonomously, such as through implicit voting schemes that assess the frequency of assertions for a particular knowledge claim.

It is, however, important to keep in mind the advice of Keeler et al. (2011), who have observed that "claiming truth by simply repeating an assertion… is a fallacy in classical logic theory". For this reason, the inductive and deductive reasoning faculties of the described Social Learning System are designed in a spirit similar to that of the *Revelator* game, also introduced by Keeler et al., which has been described by the authors as a game of complex adaptive reasoning designed to evaluate truth from strategically reasoning through logically related conjectures that are bound to existing evidence.

Clearly, knowledge retention is also essential for the effective operation of a KM system operating within this application space, where in each instance of an information extraction operation, there is the potential for new data to be harvested for the collective benefit of all future instances of information retrieval and presentation tasks. This, coupled with the transactional nature of an SLS, in some ways resembles the behavior of a collective or swarm intelligence system, where swarm intelligence – which is based on observations of life in various animal and insect kingdoms – presupposes that “rich behavior of [a collective] arises not from the sophistication of any individual entity in the colony, but from the interaction among those units” (Wolpert and Tumer, 2008).

Social Curation. The curation component of Social Learning Systems stands as yet another means of ensuring information reliability beyond the reach of automated consensus tasks. In their working paper, Malone et al. (2009) observed that “reliance on the crowd gene is a central feature of Web enabled collective intelligence systems”, and that novel, emergent collective intelligence systems tend to lean heavily on the passions and pride of their users (the “love” and “glory” genes, as they're referred to in the cited work). This is symbolic of a larger revolution that has occurred along with social computing, where users themselves form a new and dynamical component of the system, orchestrating its macro-level functionality and behavior through many micro-level interactions. As Haythornthwaite (2009) observes, “while we’ve been grappling with the question of how to gain strong, long-term, high overhead commitment to knowledge communities, another form of collaborative activity has arisen premised on exactly the opposite set of principles – weak, short-term, low overhead contributions to knowledge”. This, the author describes, is the premise for crowdsourcing, an activity that operates in the larger interest of community, though often without any formal community involvement. Crowdsourcing can serve as a highly economical substitute for dedicated experts (Nickerson et al., 2009), making it particularly appealing for systems such as ours, where it is assumed that not every potential use case or application domain for such a system will permit the inclusion of a concerted body of experts.

Li et al. (2012) highlighted the value of relying upon crowds to gather information about the workings of the world and the interactions of its inhabitants, going on to note five requirements for sociocultural knowledge crowdsourcing, which align well with knowledge crowdsourcing activities in general. These are: 1) cost-effectiveness, 2) support for "natural crowd interactions", 3) allowance for situational variation in knowledge, 4) robustness (the ability to tolerate errors, ambiguity, and other noise), and 5) proactive-ness, meaning the system should continuously strive to improve results or fill existing knowledge gaps.

It is believed that, by integrating automated reasoning by way of reasoning games akin to *Revelator*, alongside crowdsourced knowledge curation efforts, the development of a Web-scale, open-domain, and socially-enabled knowledge system can occur without an overwhelming tax being levied upon the resulting information quality or the timeliness of information processing activities.

Throughout the development of the SLS described in this work, significant consideration was also given to the user experience of its contributors, so as to promote "natural crowd interactions" (one of the crowdsourcing requirements cited earlier). It was believed that this requirement could be achieved by making crowdsourcing tasks intuitive enough to not merit a formal orientation process. Fortunately, this goal is perhaps more attainable for SLS's than could be expected in related system. With Social Learning Systems, the tiered collaboration system ensures that many of the complexities of a traditional collaborative editing process are abstracted away into more generalized curation activities. The vast majority of users will never edit the raw contents of knowledge artifacts, but can instead focus on less involved content organization and certification tasks (discussed in more detail later). The remaining requirements of Li et al. of being cost-effective, robust, and proactive align with the previously stated objectives of developing systems that are transactional, efficient at scale, dynamical in nature, and that strive to ensure the integrity of the information contents they present.

3.3 Constraints and Limitations

A number of practical constraints necessarily apply to Social Learning Systems. The first is that, due to the service-oriented nature of these systems, the ideal implementation environment would likely involve the use of commodity hardware resources. With this in mind, Social Learning Systems should be designed to operate in a somewhat economical manner with respect to hardware utilization (e.g., required CPU cycles, memory and persistent storage consumption, etc.). Though it is certainly possible that these systems could be adapted to distributed or other high performance computing environments – and there may be several benefits to doing so – this objective is not explored in our present work.

Additionally, though SLS's are, by their very nature, designed to accommodate local data sources in addition to those accessed via the web and OSN application programming interfaces, the data maintained locally for the independent development effort described in this work was limited to a custom WordNet database installation, in addition to any data retained through the regular operation of the system (e.g., for the tasks of automated reasoning or answer caching).

Finally, due to the fact that the system described here was developed independent of this research work, some details of its operations are considered proprietary, and are therefore beyond the scope of discussion for this paper. For the sake of merely providing a high-level overview of how a practical SLS can function, this has not been deemed to be a significant impediment to achieving the stated objectives of our work.

4 The Social Learning System

The described SLS was developed on a platform known as *Calico*, which provides for a large variety of information retrieval and evaluation tasks. In the Calico environment, both key-value and relational data stores are available for use. For relational data storage, an SQL-based relational database management system (RDBMS) is available, while an in-memory cache is available for key-value storage, with hard drive replication to ensure data persistence. In this case, the RDBMS employed by Calico is a modified version of the MySQL database system, although any relational database system would have sufficed. Incidentally, the prevalence of MySQL in scholarly applications and the wealth of provider and community support it offers (Vicknair et al., 2010) make it highly suitable for use in Web-scale transactional knowledge systems.

Calico provides access to the WordNet database and an expansive array of library functions for interacting with Freebase and relevant OSN API's. The modified WordNet database found in Calico replaced traditional WordNet pointers with unique relationship identifiers, which could be expanded to incorporate new relationship types, so as to allow for ontology expansion as new data sources are introduced.

A specialized module within the Calico system provides essential web scraping functionality for conducting Web- and OSN-based question answering tasks pertaining to a finite set of question types (initially, these consisted of *who*, *what*, *where*, and *when* questions). The resulting data was then filtered to remove items that appeared unrelated to the given QA task, cached locally, and then linked to the WordNet data store via a unique QA relationship pointer. From there, Calico was instructed to take over subsequent tasks, including information querying and display, handling of social curation and collaborative editing functions, and so forth.

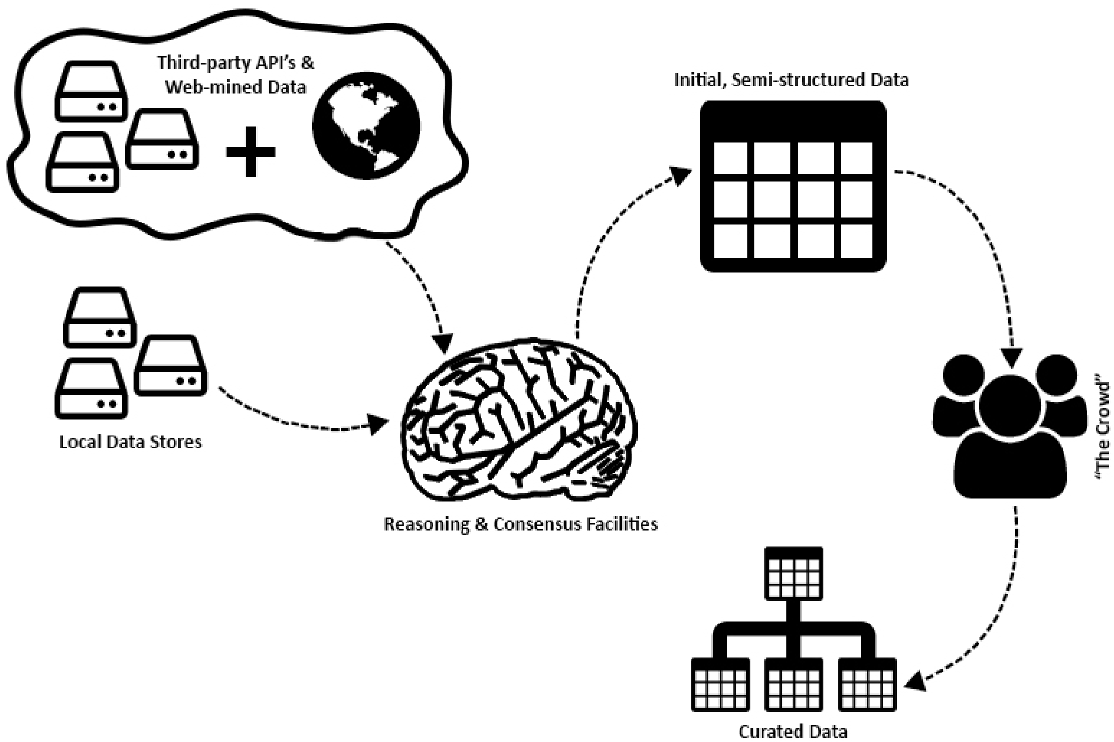


Figure 2. A basic model of the computational social learning process.

The model in Figure 2 represents a basic conceptual workflow for a Social Learning System. As we can see, SLS's rely on information contents from an array of sources, both structured and unstructured, to develop an underlying knowledge base. Information from these sources is brought into harmony through the employment of a common ontology and automated reasoning and consensus tasks designed to prevent information duplication and to estimate the reliability of information scraped from Web-based data sources. Information adapted to a common knowledge representation is then amenable to inclusion in a common ontology, where it can be further refined through social curation and collaborative editing activities. Curator contributions can then be incorporated into knowledge outputs throughout future information retrieval tasks. At the time of display, additional consensus tasks are performed to assess the perceived quality of these curator changes, in order to ensure the integrity of the underlying ontology.

As with most models, this visual oversimplifies the technical processes underlying this functionality. Through a more focused perspective, it is possible to develop a greater appreciation of the internal mechanisms of a basic SLS. As such, Figure 3 presents a more detailed visual depicting the flow of SLS execution from the time of the user's initial query to the point where output has been rendered to the screen and curation tasks can be performed.

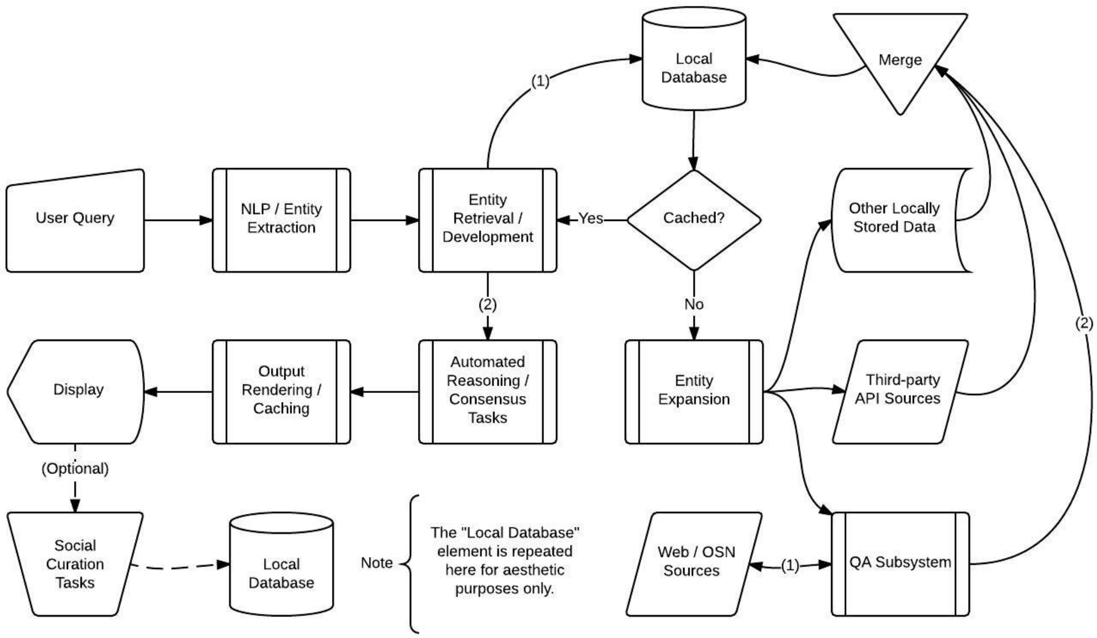


Figure 3. Sample execution flow for a Social Learning System.

As we can see from this figure, many highly specialized processes work together in responding to a user's query. Natural Language Processing (NLP) and entity extraction tasks must first resolve the focal subject of the user's query, in order that the local ontology (whose knowledge representation is made manifest in the relational database management system presented) may be consulted for possible matching entities. When data for the query subject has already been retrieved from all relevant sources, it may be returned directly to the entity retrieval and development subroutine. If not, a series of entity expansion tasks are engaged in order to retrieve and merge data from available SLS reference sources, such as Web- and OSN-mined data (via a Question Answering subsystem), other local data stores, and third-party knowledge bases.

Following this (once a complete taxonomical entity has been defined), a series of automated reasoning and consensus tasks are conducted in order to help vet information from across the various reference sources, such as through the assignment of confidence values to entity data based on source reputability or volume of factoid occurrences across sources. From there, output can be rendered into a faceted display buffer whose contents are amenable to caching, before ultimately being presented to the user. Finally, the user may opt to engage in social curation activities – or, in the case of more privileged editors, direct content revisions – which can themselves be merged into the local ontology for later assessment by the appropriate automated reasoning and consensus tasks. Hence, the process can be said to be cyclic and evolutionary in nature, as knowledge entities are constantly refined or expanded based upon the availability of source information and data generated by way of various user curation activities.

In the subsections that follow, we explore a selection of the components that have been discussed heretofore in a general sense more specifically.

4.1 Web-based and Social Question Answering

Web-based question answering systems are built from the idea that not all useful knowledge can be compressed into a single, pre-existing data store, regardless of its size or sophistication. Intellectual landscapes and the environments of our real world change perpetually, and with them, the state and availability of knowledge changes, as well. Gordon et al. (2010) noted that "the creation of intelligent artifacts that can achieve human-level performance at problems like question-answering ultimately depends on the availability of considerable knowledge". SLS's employ web- and OSN-based question answering subsystems to tap into the dynamism of real world information landscapes at scale, and to avoid information decay as the state of knowledge about the world continues to evolve.

QA systems, at the most fundamental level, exist to provide explicit, granular responses to natural language questions. In contemporary times, these systems have been instilled with the potential to address open domain questions unbound to a specific area of interest or expertise. These systems have additionally been equipped with perpetually up-to-date source contents, due to their reliance upon the World Wide Web as a first-class source of information (Kwok et al., 2001). This, in itself, makes question answering a suitable party to experiments involving Social Learning Systems, as well as an appealing reference discipline for SLS development. Through the involvement of SLS's in the advancement of Web-based, open domain QA research, we may explore the possibility of more highly sophisticated automated reasoning agents, the likes of which a great deal of early artificial intelligence researchers (and a good many works of science fiction) have led us to believe should be conversant in widely varied intellectual discourse, a task which has so far proven difficult to implement through the employment of traditional, offline corpora.

Web-based question answering has been at the helm of design decisions for QA projects for more than a decade (see, e.g., Brill et al., 2001), and is well-established as a viable means for conducting open domain answer resolution tasks. The development efforts discussed here, like the many others that came before, relied on existing web search technologies for the purpose of answer retrieval. In this arrangement, questions are transformed using a variety of rules to form one or more answer templates that can be executed as search queries to locate documents related to a particular inquiry. *Query transformations* such as these have been demonstrated in the existing literature (Brill et al., 2001), although specific query transformation and answer extraction implementations may differ considerably from one project to another.

In answer extraction, inference rules are sometimes employed to infer an answer based on semantically related, but linguistically divergent expressions. For example, the statement "Mars, Incorporated manufactures Kraft products" may be used to infer that "Kraft is a brand of Mars, Incorporated" (an inference that can, at times, lead to false conclusions). The development of these rules, however, is "extremely laborious" and "inherently difficult", due to the limitations of human rule generators (Lin and Pantel, 2001). Processing these rules is also highly computationally taxing on computer systems; even more so as the rules become more elaborate in nature.

Related work has attempted to sidestep the necessity of human rule generators by algorithmically learning surface text patterns for augmenting predefined answer extraction rules via the use of machine learning (as one example, see Ravichandran and Hovy, 2002). Again, though, the computational demands of such systems often exceed what is technically feasible for a large-scale *transactional* information processing system, particularly a Social Learning System, where information must be brought into alignment using a common ontology across multiple, disparate information sources. The technological viability of inducing such a heavy workload on a system that must interact not only with web-based information sources, but local and unindexed external sources as well, in addition to having to respond to queries in a matter of seconds (or less, in most cases), is limited.

Further, one might recall that the design of an SLS is such that it should be conducive to proper information archival, and as Wang et al. (2010) noted, "efficiently extracting temporal facts from arbitrary natural language texts with high precision is extremely difficult if feasible at all". Taken together, these sorts of technical challenges paint a somewhat bleak picture for the development of Social Learning System QA functionality. As we've discovered, however, one of the positive aspects of working with information from a variety of disparate sources, melding it into a common ontology amenable to automated reasoning tasks, and inciting the wisdom of the crowd to help curate its contents, is the ability to take certain computational shortcuts when it makes the most sense to do so (particularly when such QA tasks represents only one component of the overall system's functionality).

Beyond their reliance upon several traditional patterns of web-based question answering, Social Learning Systems also make use of social data repositories, such as through application programming interfaces (API's) made available by social networks like Twitter and Facebook. Resources such as these tend to be less suited to direct web crawling – due in large part to a lack of proper metadata and the aggregate nature of their contents – and as such are not well-indexed by search engines. Still, their contents are worth perusing within a computational knowledge context due to their ability to 1) extend the feedback loop for adjusting confidence values of existing knowledge claims, 2) to reach more distant answer outliers in the resolution of information requests, and 3) to satisfy requests whose principal focus may involve the interpretation of informal or subjective matters. In addition, Parameswaran and Whinston (2007b) have argued that "social software sites which create knowledge by collective contributions, debate and refinement tend to generate reasonably accurate information, and often lead to better insights than academic research and expensive analyst reports". While due diligence is necessary in vetting their contents, it would be inappropriate to shun the usage of large-scale social data repositories altogether, lest we lose the benefits these sources can supply to the KM process.

4.2 Local Data Stores and the Integration of Third-party Sources

Social Learning Systems attempt to present a more holistic view of world knowledge. Web- and OSN-mined data can provide a vast array of valuable inputs into the process of classifying and making sense of the world at large, but carry with them the costs of 1) having to connect to external resources, contributing to request latency and bandwidth consumption, 2) needing to provide real-time processing of retrieved data, adding to the computational overhead of each transaction, and 3) establishing increased reliance upon third-party providers. For these reasons, Social Learning Systems must either supplement or ground their knowledge repositories in localized data stores, and should cache processed contents to avoid unnecessary duplication of effort. Additionally, SLS developers should be careful not to overlook existing structured online knowledge repositories. Though they still impose certain costs (points 1 and 3 above), coming from existing structured sources, they often require little additional processing overhead, and offer a wealth of useful information for perusal in KM systems. The SLS we present here implemented a local data store amenable to data caching, in addition to importing data from the external knowledge base, Freebase, on a per-transaction basis, utilizing API's made available by its provider.

4.3 Unifying Disparate Knowledge Sources

> "An *ontology* specifies a conceptualization of a domain in terms of concepts, attributes, and relations. The *concepts* provided model entities of

> interest in the domain. They are typically organized into a *taxonomy tree* where each node represents a concept and each concept is a specialization of its parent" (Doan et al., 2003).

While it is assumed that most readers of this work will already possess some conception of what an ontology is, the definition above and the review in the earlier portions of our work faithfully capture the essence of what one must know in order to understand their significance to the KM process. Ontologies encompass some perception of what exists in the world (Evermann, 2005). They link together concepts so as to provide context and an inherent structure to the data they contain. In short, they are the social networks of larger reality, where each node may have its own existence and characteristics of that existence, while fitting into some larger schema pertaining to what is, what was, or what we perceive to be.

OWL, the *Web Ontology Language*, is perhaps the best known general purpose ontological framework. Maintained by the World Wide Web Consortium (W3C), the latest specification of the language (OWL 2) was announced in 2009, and is available for review on the W3C website (W3C OWL Working Group, 2012). OWL has been in use for over ten years, and is well-established in the KM community. Though it is a highly expressive and structured language, and is appealing for many time-insensitive offline processing tasks, OWL is computationally expensive (Wang et al., 2010), and as Davis et al. (1993) point out, "questions about computational efficiency are inevitably central to the notion of representation". The computational costs, in addition to the storage costs imposed by its highly expressive (and to some, perhaps overly verbose) syntax, make OWL a poor candidate for use in large scale, transactional web knowledge systems. With

OWL being the most stable and widely-used ontological framework to date – and with other frameworks suffering similar drawbacks – this creates a need for an expressive means of representing knowledge of the world in large scale transactional systems.

Fortunately, though it may not be an ontological framework in the nominal sense, the relational data model proposed by E. F. Codd (1970) facilitates all of the same relationship dynamics expressed by traditional ontologies, while also being widely adopted by large scale web services such as Facebook and Wikipedia (Facebook Engineering, 2012; Vaughan-Nichols, 2012) in the form of Relational Database Management Systems (RDBMS). Social Learning Systems, having the need to model data in a relational manner that is also conducive to timely, large scale information processing, rely upon relational database systems to serve as an information storage and retrieval platform for the KM task.

While an RDBMS solves a number of problems associated with adopting an ontological framework at scale, it lacks a highly crucial component for knowledge representation in ontologies: a formal taxonomical schema. Incidentally, developing a formal taxonomy of all things knowable, from the most conceptual to the most material, is not trivial (and is a task that would, naturally, be well beyond the scope of the present work). However, as Lee et al. (2011) observed, "it saves time and money if an existing taxonomy can be used to enrich a new taxonomy, and vice versa". With that in mind, what we might ask for instead of a taxonomy of *all things* is a general taxonomy of *many things*, and one that is acquiescent to change. Even still, this would be asking a lot, were one to avoid looking to the available SLS reference disciplines for inspiration. As such, in this case, a seemingly unlikely candidate was found in the domain of lexicography: the WordNet lexical database.

Most who are ill-acquainted with Princeton's WordNet database, or who have had experience using it at only in superficial capacity, will likely wonder what a dictionary has to do with structuring the knowable world. WordNet, however, exists somewhere outside the traditional boundaries of lexicography that form the mainstay of literary comprehension. In addition to encompassing the traditional mélange of parts of speech, word definitions, and example usage frames, WordNet offers a well-defined set of linguistic pointers which semantically link concepts to other, related entities. These relationships include the basics, such as synonyms and antonyms, but also more exotic relationship types, such as hypernyms (larger classes that something can be a type of, such as what *birds* are to *pigeons*) and meronyms (which indicate that something is a constituent of another, such as a *cap* being a part meronym of a *pen*). As a whole, WordNet contains more than 20 such linguistic pointers, providing a wide variety of relationship data for the vast amount of entities already contained within the database (see Table 1 for a larger listing of WordNet pointers that were preserved for use within the prototypal SLS). More importantly, these pointers are extensible in the sense that the WordNet database is freely available, and can be downloaded by researchers and practitioners alike, and adapted to any number of new uses.

The first incarnation of the Social Learning System described here employed a modified version of Princeton's WordNet database. The database served as a kind of architectural glue meant to link the contents of disparate data sources together. While it may fall short of the ideal of a complete taxonomy of all human knowledge, it does allow for 1) a single point of reference for querying concepts, and 2) the ability to supplement conceptual knowledge with relationship data across concepts, for those entities already existing within the WordNet database. One traditional problem in multi-source knowledge management that this design choice helps developers to largely avoid is that of *ontology matching*, or finding "semantic mappings" between existing ontologies (Doan et al., 2004). The sheer act of employing WordNet as a central hub in a social learning context provides for a relatively painless form of *information fusion*, or "the merging of information that originates from different sources" (Dalmas and Webber, 2007). Allowing fused concept data to benefit from existing linkages in this way assists in mitigating against the difficulties (and potentiality for errors) of a more laborious and potentially divisive ontology matching specification.

Figure 4 presents a scaled relationship model for the prototypal learning system presented in this work. The central cluster – which appears somewhat like a head of dandelion seeds – represents a graph constructed of WordNet lexical pointers. The WordNet linkages are then amended with references to new data clusters, including links to mined assertions, local data entities, tagged attributes, and encyclopedic knowledge artifacts, each of which may have structures much more elaborate than the simple visual metaphor presented here.

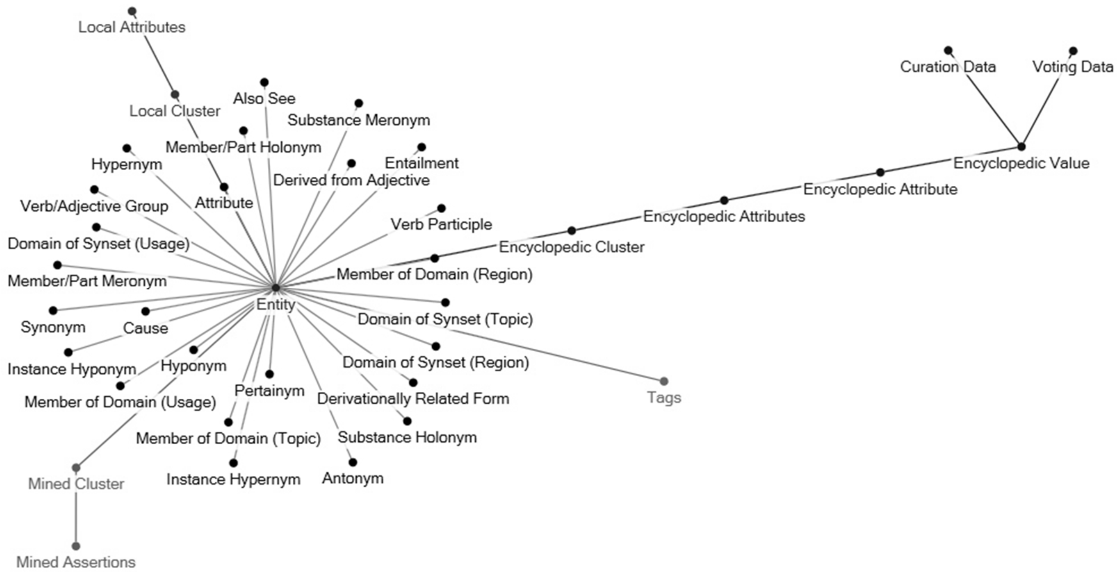


Figure 4. Scaled relationship model of a prototypal SLS.

4.4 Sample SLS Entity Relationship Diagram

Figure 5 depicts an abridged entity relationship diagram (ERD) that visualizes how the raw data of an SLS ontology can be represented within a relational database management system. It is important to note, however, that many alternative representations may exist that are just as suitable (or even superior, for that matter) for the purposes of SLS knowledge representation.

Within the diagram, we find a variety of “one-to-many” data relationships which exist to form the inherent taxonomical structure of the ontology, such as those linking entities to their WordNet sense definitions, those that link entities to one another by way of various relationship type identifiers, and those which identify attributes of entities and their corresponding values. Also of note is that many of these relationships may be ascribed a data source, which is useful for both good bookkeeping practices, as well as for conducting consensus tasks across available sources. The interpretation of the remainder of the diagram can be performed by following standard ERD practices, and is left as an exercise for the reader.

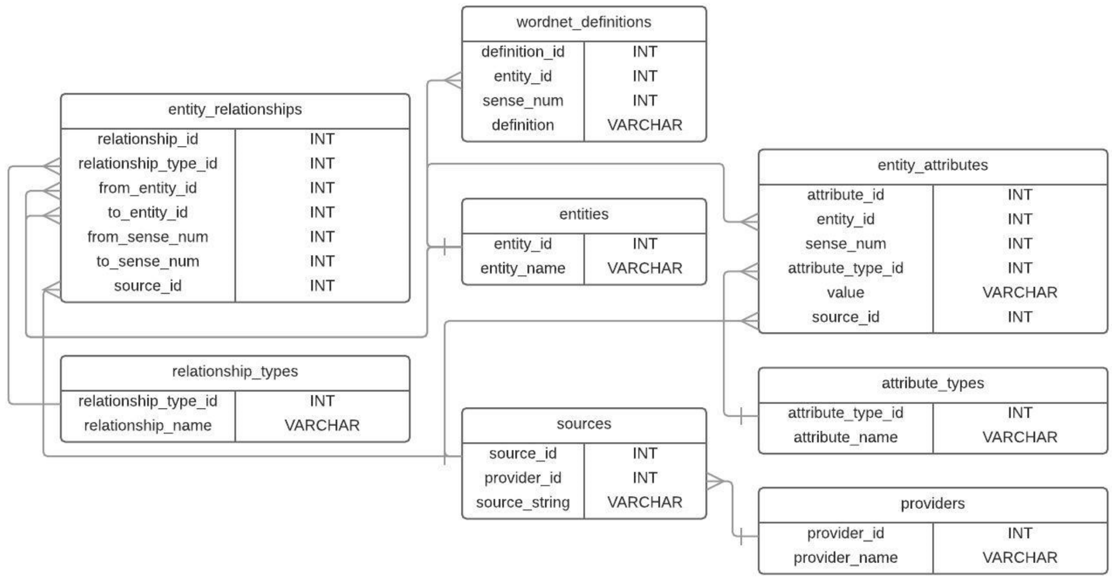


Figure 5. Example SLS entity relationship diagram.

4.5 Social Curation

In recent times, crowdsourcing has proven itself an effective means of achieving a wide variety of goals. In one study, Munro et al. (2010) were able to reproduce several classical linguistic studies involving language processing and linguistic theory using crowdsourced data in a way that proved more convenient, economical, and expeditious than the methods originally employed in the referenced studies. In a more related context, a study by Gordon et al. (2010) demonstrated that crowdsourced commonsense knowledge evaluations could correlate strongly with the assessments of AI experts within the context of an open knowledge extraction technology. This bodes well for the form of social curation employed in Social Learning Systems, as it captures not only the dynamic of individuals being willing to participate in crowdsourced knowledge projects, but also their potential to work effectively and with a duty of care. Of course, this will not hold true for every contributor in a crowdsourced effort, so effective vetting mechanisms must be in place to help offset the deleterious effects of abuse.

Social curation, as it is exists within Social Learning Systems, relies on the idea that simple actions at a micro level can have a dramatic impact at the macro level. Alternatively, as Mataric (1993) put it, “interactions between individual agents need not be complex to produce complex global consequences”. Much like the collaborative editing nature of Wikipedia, a few simple revisions or contributions might not matter in light of the massive amount of content available at a global scale, but when exponentiated by the crowd, the effects become capable of touching even the farthest reaches of the overarching knowledge infrastructure. This bears semblance to the concept of self-organization, wherein global results are orchestrated via the operationalization of local information (Tarasewich and McMullen, 2002).

In a related sense, the social curation efforts of an SLS can be thought of as a model of distributed intelligence, as well, where entities work together “to reason, plan, solve problems, think abstractly, comprehend ideas and language, and learn” (Parker, 2008). These actions become effective at larger scales through a system of collective interactions, where users share goals and the actions of one are “beneficial to their teammates” (where in this case, teammates could refer to fellow curators, editors, and end users of the system, alike).

The implementation of the social curation construct consisted of a system of annotated indicators. In this system, all knowledge artifacts (relationships, entities, or attributes) can be *flagged*, where such flags may take on a variety of meanings. Two examples of this are the *confirmed* or *refuted* flags, where a “confirmed” indicator may be used to suggest that an item is verifiable via corroborating external sources, while a "refuted" indicator suggests that an item is contestable, such as being invalid for a given context or more generally falsifiable (e.g., a user may refute a relationship to indicate that a concept does not belong within a larger class of concepts). Each flag is also annotatable; the user submitting the indicator may supplement their submission with additional explanatory details. Further, "confirmed" and "refuted" flags are able to be ascribed a source (via URL) for verifying their asserted claims.

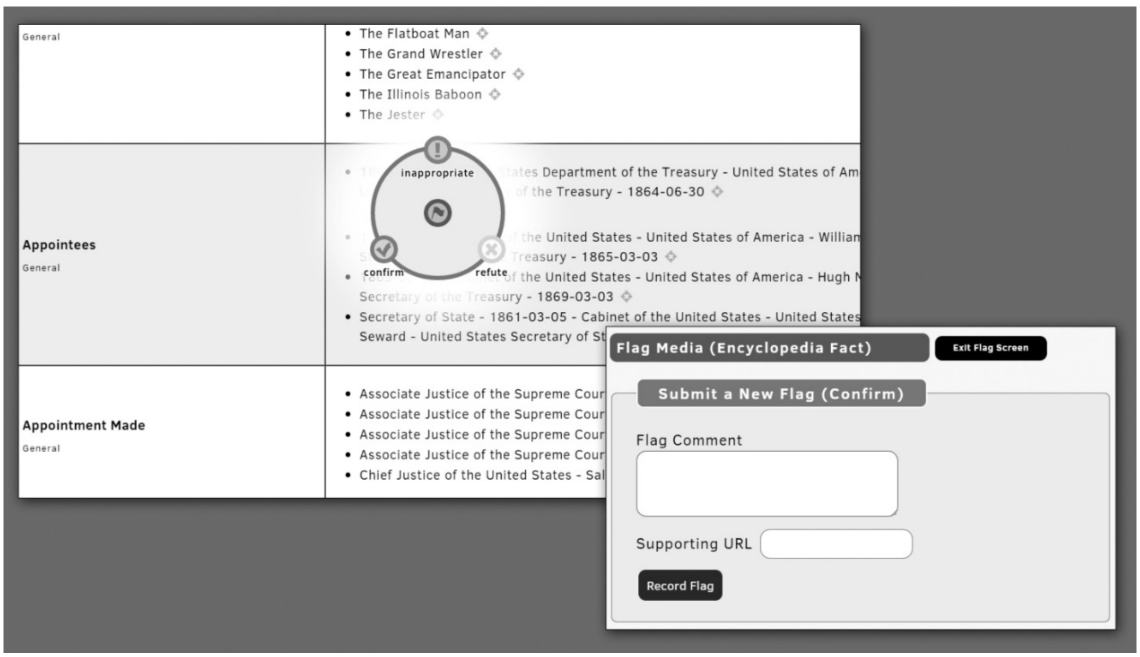


Figure 6. Indicating a flag type for submission in an SLS.

This flagging system provides users with a quick and convenient way to curate SLS knowledge contents, though taken alone, it would result in a considerable burden for system administrators, who could quickly become overwhelmed by the task of evaluating submitted flags and responding to each claim, individually. Therein lies an opportunity for the automated response mechanisms of an SLS, which should be capable of mechanistically performing the following actions: 1) displaying visual feedback for "confirmed" knowledge artifacts, and 2) intelligently deescalating the prominence of "refuted" contents. Each of these tasks should occur only once a certain minimum threshold of indicators has been surpassed. To help prevent the system from being gamed by potential abusers, user IP addresses and (if applicable) associated account identification information should be logged alongside all flag submissions. In this way, duplicate flag submissions and submission spamming attempts can be mediated.

Parameswaran and Whinston (2007a) have observed that "highly dynamic and decentralized communities engaging in grassroots innovation lead to significant unpredictability in the system". For this reason, coordination of knowledge tasks is crucial for users contributing to the curation process. Further, coordinating the collaboration process becomes particularly important as the group of contributors grows in size (Kittur and Kraut, 2008). One way in which Calico attempts to coordinate efforts of knowledge workers is via the employment of a collaborative tagging system, where outside of existing relationship declarations, users may directly tag knowledge artifacts with relevant classifying information. These tags then allow users to quickly locate related contents that have been associated with the same or similar tags. This is valuable for users with expert knowledge in a particular domain, who may wish to quickly access similar items relating to their areas of interest for curation purposes. The use of a collaborative tagging system (or a *folksonomy*, as it's been referred to in other contexts) for knowledge management has been cited to result in lower costs to the user (in contrast to more complicated, hierarchical taxonomies), increased flexibility or dynamism of categorizations, and a more "democratic" means of meta data generation (Wu et al., 2006).

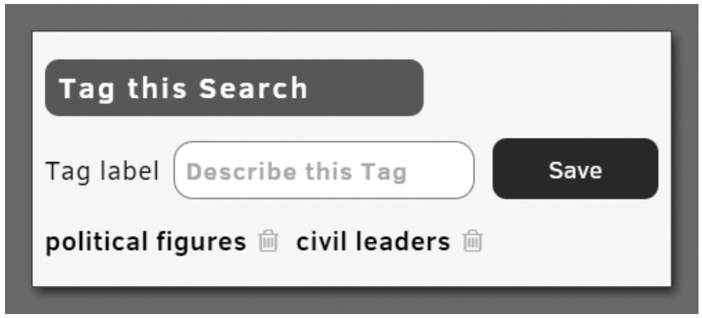


Figure 7. Submitting a tag in Calico.

Collaborative tagging, in conjunction with the content flagging system described earlier, contributes to a larger democratic means of social curation and knowledge content management. As Brabham (2008) admonishes, however, collaborative projects should be cautious in "assuming that ideas emerging from the crowd… represent an ascendance of the superior idea through democratic process" (Brabham, 2008). Certain biases present themselves in all manner of human endeavors, and crowd-sourced KM efforts are no exception. Individuals and larger groups are subject to cultural and demographic predispositions, cognitive biases, and no small amount of demagoguery. A good summary of the various heuristics and biases individuals may find themselves subjected to is presented in Table 1 of Schwenk (1988), reproduced in Figure 8 of this text. While solving for this problem entirely may not be possible, it is possible to vet knowledge claims via links to reputable sources (such as those that are allowed when flagging confirmed or refuted contents), as well as to provide more expansive feedback loops, incorporating the views of a larger portion of the overall content audience, and therefore making the existing "democratic process" somewhat less selective in nature.

Table I. Selected heuristics and biases

Bias	Effects
(1) Availability	(1) Judgements of probability of easily-recalled events distorted.
(2) Selective perception	(2) Expectations may bias observations of variables relevant to strategy.
(3) Illusory correlation	(3) Encourages belief that unrelated variables are correlated.
(4) Conservatism	(4) Failure sufficiently to revise forecasts based on new information.
(5) Law of *small* numbers	(5) Overestimation of the degree to which small samples are representative of populations.
(6) Regression bias	(6) Failure to allow for regression to the mean.
(7) Wishful thinking	(7) Probability of desired outcomes judged to be inappropriately high.
(8) Illusion of control	(8) Overestimation of personal control over outcomes.
(9) Logical reconstruction	(9) 'Logical' reconstruction of events which cannot be accurately recalled.
(10) Hindsight bias	(10) Overestimation of predictability of past events.

Figure 8. Select Heuristics and Biases, reproduced from Schwenk (1988).

Following from this latter point, the Calico system incorporated a built-in, context-free voting mechanism for providing feedback with no requirement for additional explanatory details (i.e., users were not required – and in fact, were entirely unable – to provide annotations or source material references, as they would when flagging contents). This system was enabled for all knowledge contents, allowing even casual users without a highly vested interest in the knowledge artifacts to quickly express positive or negative reactions that were evoked in response to a particular SLS assertion.

Taken together, these three mechanisms (content flagging, collaborative tagging, and context-free voting) provide a means of social curation that is convenient and powerful, and which importantly never requires users to actually invest the time required to draft original contents on their own, due to the automated sourcing of information from existing repositories. One benefit that this provides is that curation efforts are able to take place much more quickly, as these functions occur at a level immediately overlaying the content generation process.

These social curation faculties may also serve as triggers to engage in further automated or manual processes, such as automated content hiding for refuted or low-quality factoids, or signaling to a group of more highly privileged editors that a particular knowledge entity needs more material revisions. In this way, a tiered contribution system is formed, wherein casual users can easily (without substantial technical know-how) curate knowledge contents, while those in editorial roles can partake in more traditional collaborative editing activities.

4.6 Social Trust

In social computing, it can be said that users themselves represent the foundation of a social technology infrastructure. As such, in a KM context, we must acknowledge that "the reliability of the user providing the information is as important as the information they provide" (O'Donovan, 2009). This observation highlights the need for a system of social trust, wherein the KM system is engineered with an appropriate means to vouch for the integrity of the various participants of the KM process. While it has been observed that merely participating in a community of shared ideals and common goals results in the development of a basic level of trust (Bialski and Batorski, 2010), due diligence must be exercised to ensure the trust bestowed has not been ill-placed.

Implementing the capacity to vouch for a user's trustworthiness typically relies on the notion of a trust metric, consisting of "the different computations and communications which are carried out by the trustor (and his/her network) to compute a trust value in the trustee" (Seigneur, 2009). In a KM context, one perhaps obvious means of establishing this trust value is via an examination of the aggregate fidelity of a user's knowledge contributions. If a user consistently confirms facts or prompts structural revisions to the knowledge graph that are supported by the community at large, it can be roughly assumed that the user is a trustworthy participant within the KM function. The antithesis of this would be the case where a user supports changes that are at odds with the larger knowledge community, or if they exhibit questionable feedback patterns (such as submitting an excessive number of a specific type of flag with no explanatory details, or an excessively high volume of negative votes).

The SLS described here implemented a system of implicit trust-granting, with a variable trust metric that was adjusted automatically based on factors such as a user's aggregate disconformity with the larger contributor community, the detection of irregular patterns of activity, and the relative number of curatorial actions reversed by those in more privileged editorial roles, who themselves are promoted from a pool of established, highly active contributors with sufficiently positive trust metrics.

4.7 Automated Reasoning and Consensus Tasks

Reliance upon the crowd for driving large-scale cooperative efforts has already proven to be an economical means of achieving a variety of goals. With that said, in a social curation context, contributors cannot be expected to work very effectively without an established and semi-organized baseline from which to conduct their work. More to the point, it doesn't much matter if a user wishes to curate an entry on *dogs* if, when extracting and coalescing source materials, the system determined that a dog was a type of fruit. In short, the system must be equipped with a means of establishing a minimal canvas for curators to be able to effectively engage in their art.

It has been mentioned that the core of the SLS knowledge graph rests on a modified version of the WordNet database. Presiding over this, at the data retrieval level, the system employs a set of basic reasoning tasks designed to 1) reduce factoid duplication across sources through similarity analysis, and 2) in the case of Web- and OSN-mined data, establish consensus for extracted factoids across information sources, which is accomplished by tracking the number of occurrences of sources supporting a particular mined assertion (e.g., the number of users on a given social platform repeating a claim, or the number of distinct web sources publishing a related statement pertaining to the subject of interest).

Consensus tasks also find a home in the curation stage of the social learning process, where confirmation, refutation, and other indicators levied against knowledge artifacts are weighted together to determine automated responses, such as downgrading the priority of a factoid or removing erroneous relationships. These automated tasks, coupled with crowdsourced, manual curation, minimize the need for administrative oversight for the Social Learning System, simultaneously reducing operational costs for organizations implementing SLS's.

4.8 Querying the SLS

When interacting with the SLS, the Calico system employed a natural language querying model and rudimentary answer resolution faculties to allow for more fine-grained inspections of knowledge contents, such as explicit attribute-level queries for known concepts (e.g., the birthdate of a famous historical figure), comparisons across concepts (for example, contrasting properties of related concepts, such as Google and Microsoft, both large, multinational, and primarily technology-oriented companies), and eventually, inferences or deductions throughout the concept taxonomy (such as to infer missing properties for a member of a class based on the constitution of the class as a whole). This last part is highly significant, as such reasoning across concepts can be made vastly more efficient in a general purpose knowledge representation, due to the lack of mappings across taxonomies or other artificial bridging mechanisms which typically incur significant costs at either the data storage or application level.

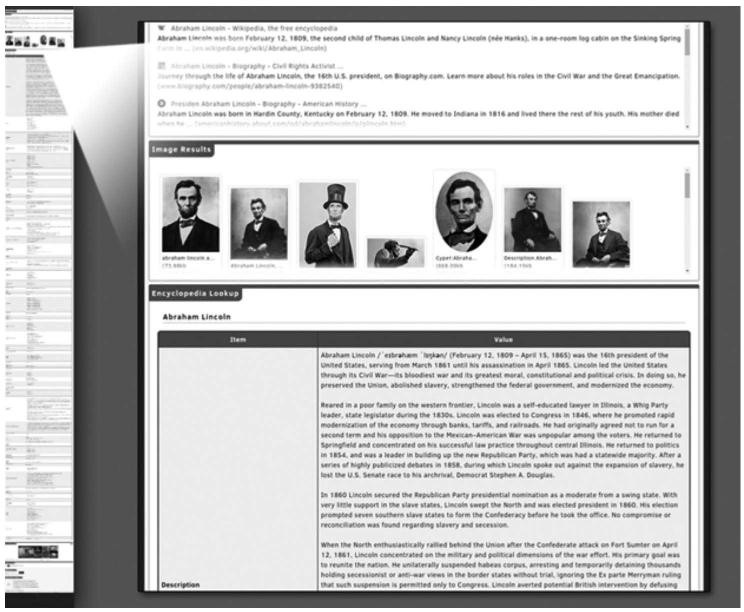


Figure 9. Faceted content display by an SLS.

Following a successful query, output for the user is rendered in a series of panels, each featuring a different facet of the available media relating to the particular knowledge entity being explored. While source materials may be unstructured or loosely structured in nature, the output of the Social Learning System itself will always be highly structured, in accordance with the SLS's own internal representation of knowledge artifacts. Within the output, an advanced Social Learning System may incorporate not only textual information, but also video, images, or even audio files relating to the topic of interest, in order to provide a more holistic view of the subject at hand. This results in a potentially significant amount of content being presented to the user at once, and as a result, SLS developers should pay close attention to the manner in which contents are exposed to end users, to ensure that the mode of expression is conducive to the end-user's own information acquisition and decision-making processes. For more detailed coverage of this topic, readers are encouraged to review *Appendix C: Information Presentation and Decision-Making*. Readers may also be interesting in reviewing *Appendix F: Additional Screenshots*, for additional screenshots of the prototypal SLS discussed here.

5 Evaluation of the SLS Artifact

At the data storage level, the Social Learning System described here relied upon a MySQL-like database management system as a means for storing and retrieving knowledge graph data. Each knowledge graph entity (subject) had potential prepopulated fields contributed by WordNet and other local data sources. However, as this information was somewhat limited, for our purposes, we assume that a basic knowledge entity is only established for practical purposes after it has had values contributed to it from external sources. In the case of the described SLS, perhaps the most notable are the relevant encyclopedic data API's, which provide a wealth of diverse information suitable for sophisticated reasoning tasks, the likes of which SLS's have been specifically designed to accommodate.

Due to its transactional nature, the SLS did not actively harvest new information until being compelled to do so, at which point a series of library functions provided by the underlying application framework, Calico, initiated API calls to both Wikipedia (a predominantly unstructured data source) and Freebase (a predominantly structured data source), querying for related subjects and using specially designed parsers to extract data relating to the most likely subject match for each user query. (In this particular application, a proprietary Wikipedia parser was used, though for future development initiatives, it should be noted that the DBpedia and Wikidata projects aim to provide much the same functionality.)

Because SLS's are intended to be capable of engaging in transactional knowledge tasks at scale – while employing commodity hardware – our evaluation in this section is focused solely on the measured computational efficiency of an SLS prototype. In each evaluation, a number (n=100) of ontological development tasks are performed, and the average time to completion is assessed at a granular level. Further, various forms of caching are introduced – and additional metrics are obtained – to measure the impact of different caching schemes on SLS efficiency.

An initial performance evaluation was conducted in a highly restrictive demonstration environment boasting a somewhat lackluster array of system resources, even when compared to low-end commodity web servers. This environment featured a Windows 8.1 Pro (64-bit) operating system with available resources consisting of a dual-core Intel Core i5-2467M with a base frequency of 1.6 GHz, 4GB DDR3 RAM, and a 128GB SSD drive.

When simultaneously developing a basic knowledge entity for the first time (without previous API calls or query caches to rely on) and rendering output to a user in an active system environment where other working processes interacted with both the database and web server, a typical total script execution concluded in about 5 seconds, with ~98.6% of that time devoted to general knowledge tasks and about 70.1% of the total execution time devoted specifically to SLS-related encyclopedic data retrieval, parsing, and consensus tasks. Of the portion of execution time devoted strictly to encyclopedic data tasks, ~73% of the time was focused on working with data from Freebase (with the bulk of that time – 77.9% – spent waiting for information to be received via the remote API). About 21.3% of encyclopedic data execution time was allocated to Wikipedia retrieval, parsing, and storage tasks, wherein about 83% of that time was once again spent waiting for data from third-party API calls. This idle time is of particular note, due to it being unavoidable in this type of transaction on a system with very limited local storage capabilities that would prohibit bulk data downloads. Further, as it accounts for the vast majority of script execution time, it helps to highlight the relative efficiency of parsing, consensus, and storage tasks, which themselves conclude in a small fraction of the original time spent waiting for data to be received from third-party sources.

Though these results are somewhat optimistic for SLS's, when we consider subsequent graph queries that implement caching (wherein third-party API connections are no longer necessary), the computational burden decreases tremendously. In a caching scheme strictly implementing data (and not output) caching, average execution time fell to about .4 seconds for knowledge management-related tasks, including those that are not SLS-specific. Meanwhile, encyclopedic data retrieval from cache and output rendering accounted for approximately 65.5% of total execution time, a proportional 5% decrease on an already less taxing transaction.

With more aggressive caching, wherein both data and output rendering are cached for all knowledge entities, average execution time for all KM-related tasks fell to under

200 milliseconds (~63.5% of total script execution time, while cache retrieval and output rendering accounted for less than one half of a percent of total execution time).

In systems featuring more capable central processing units, more expansive memory and persistent storage space, and faster persistent storage or underlying system bus transfer rates, computation time could be reduced not only due to inherent system improvements, but also due to additional opportunities that expanded system resources would allow. As one example, both Freebase and Wikipedia, which were queried remotely in the application described here, offer data dumps for compressed, bulk downloads of their data contents at regular intervals. Integrating these bulk downloads within systems where such is feasible would greatly reduce or eliminate the idle time that accounts for the majority of the script execution time encountered in the non-cached SLS lookups. Additionally, where this application relies on persistent storage for data and output caching (albeit via an SSD hard disk), additional performance could be realized through the adoption of an in-memory caching scheme, for systems with sufficient RAM.

Example factoids from the prototypal SLS for the query "Albert Einstein" are included in Table 2 of *Appendix E*, in order to give an idea of the variety of information output that can be expected from such a system. Keep in mind, however, that the sample output has been obtained from a non-public system, and therefore has not, to this point, been curated or edited in any way.

6 Future Research Directions

By this time, it should be clear that Social Learning Systems possess outstanding potential for adoption within a variety of KM and artificial intelligence contexts. Even so, this work has only just begun to examine the capabilities of this new class of systems.

Opportunities for future research might include demonstrating practical applications of Social Learning Systems in various domains, integrating deeper linguistic analysis faculties into the Web- and social-mining functions to extract factoid-level data as opposed to general assertions (thus creating more structured data amenable to future reasoning tasks), performing graph analytic evaluations upon a populated knowledge graph for entity-level deduction and inference tasks, and more.

Alternatively, at a higher level, researchers may evaluate Social Learning Systems in light of different theoretical or methodological lenses, perhaps grounding SLS application development or modeling initiatives in theoretical frameworks more appropriate for specific usage scenarios. Others, still, may examine ways in which SLS's can be viewed in light of more general information systems-centric conceptual frameworks, such as the nascent Philosophy of Information (discussed in *Appendix D*), which help bridge the divide between various computer science cognitive disciplines.

Another interesting future research direction could involve analyzing collaboration patterns with regards to Social Learning Systems, which in addition to traditional collaborative editing functionality, introduce new modes of digital social curation. In their work, Kittur et al. (2007) observed that large scale online collaboration projects tend to be driven predominantly by a small number of prolific early users. It is these users who help to define the utility of the nascent system and pave way to more mainstream adoption, at which time more generalized contributions expand in relation to those of the early contributors to the system. What might this distribution and evolution of activity look like in a large-scale Social Learning System? Further, what might the make-up of resulting contributor base look like? Wikimedia (Wikipedia's parent organization) user statistics shine light on some of the consequences of a large-scale KM system adopting a purely editorial model. According to the published figures, only about 0.02 to 0.03 percent of all visitors to Wikipedia are active contributors to the site. Of these, fewer than 15% are female, around 70% are single, and fewer than 20% have children (Wikimedia Users, n.d.). It would be interesting to see if these same patterns manifest in KM systems with lower barriers to entry for contributors. As it stands, many demographics seem alienated from the collaboration processes of related systems.

Lastly, there remains an open research question regarding how best to ascertain the accuracy of results contained within collaboratively edited knowledge projects. In their write-up on the architecture of the Never-Ending Language Learner (NELL) project, Carlson et al. (2010) employed manual precision evaluation by humans. Manual validation was also performed in the related work of Vinyals and Le (2015), for a machine learning project at Google involving the development of an intelligent chat bot. A large number of additional studies have been conducted employing manual validation in an effort to ascertain the accuracy of Wikipedia (Reliability of Wikipedia, n.d.). Clearly, existing evaluation practices are costly and difficult to manage at scale. Future research that seeks to establish more automated means of estimating result accuracy in collaboratively edited knowledge collections may prove useful in this light.

7 Key Contributions

Prior to the conclusion of this work, it is worthwhile to reflect on some of the key contributions that were made as a result of this exposition on the workings of a prototypal Social Learning System. The most notable among these, in our view, are presented in the numbered list that follows:

1. The work has contributed a model for defining generalizable ontologies in a highly pragmatic way, using relational data structures to allow for efficient information storage and retrieval for knowledge tasks at scale. As discussed, the limitations of existing ontological frameworks would have proved prohibitive for the development of Social Learning Systems and the performance of computational knowledge tasks at scale. By employing a relational data model and relying upon existing relationship data made available via Princeton's WordNet lexicographical database, we were able to develop a general purpose ontology that is both highly expressive and computationally efficient.
2. We have developed a unification framework for allowing ontological development from disparate, unlike data sources, including unstructured web and social sources, semi-structured knowledge sources like Wikipedia, and structured data repositories, such as Freebase and WordNet.
3. Lastly, our work has defined a novel, tiered collaborative editing structure that, in addition to traditional collaborative editing features, allows for less abstruse knowledge curation tasks involving simple crowdsourced feedback mechanisms. This form of social curation, in the SLS context, allows users to become active participants in the collaborative editing process, without a significant time investment or involvement in a more formal system orientation.

8 Conclusion

Nguyen (2015) stated that "the *technology* of machine learning is giving us new ways to think about the *science* of human thought ... and imagination". Perhaps only by advancing the state of the art, both from a practical and theoretical perspective, can we hope to push closer to the ever lofty objective of constructing a more perfect looking glass for peering into our own minds and exploring their inner mechanisms. Forbus (2012), for one, has attested that we may only ever truly attain an understanding of the complexities of developing minds "by using components to build integrated cognitive systems… And yet," he continues, "today, almost all work in artificial intelligence falls into the brick-making mold". The present work in describing a model for Social Learning Systems is just one attempt at escaping that mold.

With an emphasis on pragmatism, Parameswaran and Whinston (2007b) proclaimed that, "from the viewpoint of a user, what matters is the value of knowledge created, and not how it was created". Though this may be true, the process of knowledge creation is inherently entwined with a system's latent value generation. The less rigorous and comprehensive the knowledge creation process, the less value will be obtainable from that process. To this end, Valiant (1984) suggested that the design of a learning machine should include each of the following properties: 1) that they possess the ability to demonstrably learn "whole classes of concepts" that can be characterized; 2) that the concepts they learn are "appropriate and nontrivial for general-purpose knowledge"; and 3) that the process of deducing what is learned involves a "feasible (i.e., polynomial) number of steps". In this work, we have described a class of systems that ostensibly satisfies each of these criteria in a matter that is suitable for large scale web interactions.

Social Learning Systems, as presented throughout our research, have been envisioned as dynamical knowledge management systems that integrate Web- and OSN-mined data alongside data retrieved from third-party API's and local data stores, in conjunction with social curation, collaborative editing, and automated reasoning faculties, in order to competently assess and provide desired information spanning an unrestricted array of knowledge domains. Built on the shoulders of giants such as Wikipedia, Freebase, and WordNet, an independent SLS development effort was described that presented a unified ontological framework with a degree of computational efficiency conducive to handling open domain interactions at scale, while being effectively deployable on commodity hardware resources.

The challenges in developing such a system were substantial; the rewards for successfully overcoming these challenges, however, are even more substantial. Social Learning Systems expand the horizon of open domain KM systems and automated reasoning, promote an ontological model of general applicability conducive to transactional processing, attempt to decrease the naive error present in many Web-enabled knowledge systems via tiered social curation and editorial faculties, and reduce the intractability of more traditionally non-computable knowledge problems that exist when working with "fuzzy" or unstructured data.

By following a real world development effort for constructing a Social Learning System, the practical feasibility of this class of systems has been established. This has elevated SLS's above the realm of hazy and untested theoretical constructs, while also opening the doors to a wide range of experimental analyses that may be conducted upon future SLS variants implemented within production systems. Such analyses could provide valuable empirical data to scholars and practitioners alike who may have interest in these and related fields, while perhaps also unveiling new functional potentials that can be made manifest in the increasingly ambitious knowledge systems of the future.

References

Arasu, A., & Garcia-Molina, H. (2003). Extracting Structured Data from Web Pages. *Proceedings of the 2003 SIGMOD International Conference on Management of Data*, 337-348.

Arrow, K. J. (1994). Methodological Individualism and Social Knowledge. *The American Economic Review*, *84*(2), 1-9.

Baldwin, D., & Yadav, S. B. (1995). The Process of Research Investigations in Artificial Intelligence - An Unified View. *IEEE Transactions on Systems, Man, and Cybernetics*, *25*(5), 852-861.

Bandura, A. (1971). *Social Learning Theory*. General Learning Corporation.

Bateman, J., Kamps, T., Kleinz, J., & Reichenberger, K. (2001). Towards Constructive Text, Diagram, and Layout Generation for Information Presentation. *Computational Linguistics*, *27*(3), 409-449.

Bechtel, W. (2009). Constructing a Philosophy of Science of Cognitive Science. *Topics in Cognitive Science*, *1*(3), 548-569. doi:10.1111/j.1756-8765.2009.01039.x

Benjamins, V. R., Fensel, D., & Perez, A. G. (1998). Knowledge Management through Ontologies. *Proceedings of the 2nd International Conference on Practical Aspects of Knowledge Management (PAKM '98)*.

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research*, *25*(3), 187-217.

Bialski, P., & Batorski, D. (2010). From Online Familiarity to Offline Trust: How a Virtual Community Creates Familiarity and Trust between Strangers. In P. Zaphiris & C. S. Ang (Eds.), *Social Computing and Virtual Communities* (pp. 179-204). CRC Press.

Biehal, G., & Chakravarti, D. (1982). Information-Presentation Format and Learning Goals as Determinants of Consumers' Memory Retrieval and Choice Processes. *Journal of Consumer Research*, *8*(4), 431-441.

Boden, M. A. (2006). *Mind as Machine: A History of Cognitive Sciences* (Vol. 1). Oxford, England: Clarendon Press.

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge. *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, 1247-1249.

Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication, 13(1), 210-230.

Brabham, D. C. (2008). Crowdsourcing as a Model for Problem Solving. *Convergence: The International Journal of Research into New Media Technologies*, *14*(1), 75-90.

Brewster, C., & O'Hara, K. (2007). Knowledge Representation with Ontologies: Present Challenges - Future Possibilities. *International Journal of Human-Computer Studies*, *65*(7), 563-568.

Brill, E., Lin, J., Banko, M., Dumais, S., & Ng, A. (2001). Data-Intensive Question Answering. *Proceedings of the Tenth Text REtrieval Conference (TREC '01)*, 393-400.

Brook, A., & Mandik, P. (2007). The Philosophy and Neuroscience Movement. *Analyse & Kritik*, *29*, 3-23.

Buneman, P., Davidson, S. B., Fernandez, M., & Suciu, D. (1997). Adding Structure to Unstructured Data. *6th International Conference on Database Theory (ICDT), Lecture Notes in Computer Science (LNCS)*, *1186*, 336-350.

Bynum, T. W. (2010). Philosophy in the Information Age. *Metaphilosophy*, *41*(3), 420-442. doi:10.1111/j.1467-9973.2010.01651.x

Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka Jr., E. R., & Mitchell, T. M. (2010). Toward an Architecture for Never-Ending Language Learning. *In Proceedings of the Conference on Artificial Intelligence (AAAI 2010)*, 1306-1313.

Carter, M. (2007). *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence*. Edinburgh University Press.

Chau, P. Y. K., Au, G., & Tam, K. Y. (2000). Impact of Information Presentation Modes on Online Shopping: An Empirical Evaluation of a Broadband Interactive Shopping Service. *Journal of Organizational Computing and Electronic Commerce*, *10*(1), 1-22.

Chisholm, R. M. (1941). Sextus Empiricus and Modern Empiricism. *Philosophy of Science*, *8*(3), 371-384.

Christian, B. (2011). *The Most Human Human: What Talking with Computers Teaches Us About what it Means to Be Alive*. Doubleday.

Codd, E. F. (1970). A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM*, *13*(6), 377-387.

Cohen, P. R., & Howe, A. E. (1989). Toward AI Research Methodology: Three Case Studies in Evaluation. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(3), 634-646.

Coiera, E. W. (1996). Artificial Intelligence in Medicine: The Challenges Ahead. *Journal of the American Medical Informatics Association*, *3*(6), 363-366.

Cooke, L. (2005). A Visual Convergence of Print, Television, and the Internet: Charting 40 Years of Design Change in News Presentation. *New Media & Society*, *7*(1), 22-46. doi:10.1177/1461444805049141

Cooley, C. H. (1926). The Roots of Social Knowledge. *American Journal of Sociology*, *32*(1), 59-79.

Cross, N. (2001). Designerly Ways of Knowing: Design Discipline Versus Design Science. *Design Issues*, *17*(3), 49-55.

Dalmas, T., & Webber, B. (2007). Answer Comparison in Automated Question Answering. *Journal of Applied Logic*, *5*, 104-120.

Davis, R., Shrobe, H., & Szolovits, P. (1993). What Is a Knowledge Representation? *AI Magazine*, *14*(1), 17-33.

Dayhoff, J. E., & DeLeo, J. M. (2001). Artificial Neural Networks: Opening the Black Box. *Cancer*, *91*(Suppl. 8), 1615-1635.

de Bruijn, O., & Spence, R. (2000). Rapid Serial Visual Presentation: A Space-Time Trade-Off in Information Presentation. *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '00)*, 189-192.

Dewey, J. (1910). *How We Think*. Boston, MA: D.C. Heath and Company.

Doan, A., Madhavan, J., Dhamankar, R., Domingos, P., & Halevy, A. (2003). Learning to Match Ontologies on the Semantic Web. *The International Journal on Very Large Data Bases*, *12*(4), 303-319.

Doan, A., Madhavan, J., Domingos, P., & Halevy, A. (2004). Ontology Matching: A Machine Learning Approach. *Handbook on Ontologies*, 385-403.

Duh, K., Hirao, T., Kimura, A., Ishiguro, K., Iwata, T., & Yeung, C.-M. A. (2012). Creating Stories: Social Curation of Twitter Messages. *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, 447-450.

Dumais, S., Banko, M., Brill, E., Lin, J., & Ng, A. (2002). Web Question Answering: Is More Always Better? *Proceedings of the 25th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2015)*, 291-298.

Dupuy, J.-P. (2000). *The Mechanization of the Mind: On the Origins of Cognitive Science*. Princeton University Press.

Evaristo, J. R., & Karahanna, E. (1997). Is North American IS Research Different from European IS Research? *Database for Advances in Information Systems*, *28*(3), 32-43.

Evermann, J. (2005). Towards a Cognitive Foundation for Knowledge Representation. *Information Systems Journal*, *15*(2), 147-178.

Facebook Engineering. (2012, March 4). MySQL and Database Engineering: Mark Callaghan. Retrieved from https://www.facebook.com/notes/facebook-engineering/mysql-and-database-engineering-mark-callaghan/10150599729938920

Floridi, L. (2002). What is the Philosophy of Information? *Metaphilosophy*, *33*(1/2), 123-145.

Forbus, K. D. (2012). How Minds Will Be Built. *Advances in Cognitive Systems* (1), 47-58.

George, F. H. (1979). *Philosophical Foundations of Cybernetics*. Kent, Great Britain: Abacus Press.

Gordon, J., Van Durme, B., & Schubert, L. K. (2010). Evaluation of Commonsense Knowledge with Mechanical Turk. *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (CSLDAMT '10)*, 159-162.

Greenwood, D. (2007). An Overview of Neural Networks. *Behavioral Science*, *36*(1), 1-33. doi:10.1002/bs.3830360102

Guarino, N. (1995). Formal Ontology, Conceptual Analysis and Knowledge Representation. *International Journal of Human-Computer Studies*, *43*(5-6), 625-640.

Hall, R. P., & Kibler, D. F. (1985). Differing Methodological Perspectives in Artificial Intelligence Research. *The AI Magazine*, *6*(3), 166-178.

Hartley, R. (1928). Transmission of Information. *Bell System Technical Journal*, *7*(3), 535-563.

Haythornthwaite, C. (2009, May 31). *Online Knowledge Crowds and Communities* [PDF]. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/14198/Online%20Knowledge%20Crowds%20and%20Communities_Reno_2009.pdf

Hemmatazad, N. (2014). Social Computing. In M. Khosrow-Pour (Ed.), Encyclopedia of Information Science and Technology (3rd ed., pp. 6754-6761). IGI Global.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75-105.

Heylighen, F. (1999). Collective Intelligence and its Implementation on the Web: Algorithms to Develop a Collective Mental Map. *Computational & Mathematical Organization Theory*, *5*(3), 253-280.

Horn, R. E. (1999). Information Design: Emergence of a New Profession. In R. Jacobson (Ed.), *Information Design* (pp. 15-33). Cambridge, MA: The MIT Press.

Kamar, E., Hacker, S., & Horvitz, E. (2012). Combining Human and Machine Intelligence in Large-scale Crowdsourcing. *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, *1*, 467-474.

Kankanhalli, A., Tanudidjaja, F., Sutanto, J., & Tan, B. C.Y. (2003). The Role of IT in Successful Knowledge Management Initiatives. *Communications of the ACM*, *46*(9), 69-73.

Keeler, M., Johnson, J., & Majumdar, A. (2011). Crowdsourced Knowledge: Peril and Promise for Complex Knowledge Systems. *Proceedings of the Eighth International Conference on Complex Systems (ICCS 2011)*, 753-767.

Kittur, A., Chi, E., Pendleton, B. A., Suh, B., & Mytkowicz, T. (2007). Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie. 25th Annual ACM Conference on Human Factors in Computing Systems (CHI 2007).

Kittur, A., & Kraut, R. E. (2008). *Harnessing the Wisdom of Crowds in Wikipedia: Quality through Coordination* (Research Report No. 99). Human-Computer Interaction Institute.

Kwok, C., Etzioni, O., & Weld, D. S. (2001). Scaling Question Answering to the Web. *ACM Transactions on Information Systems*, *19*(3), 242-262.

Lee, C. A., Tibbo, H. R., & Schaefer, J. C. (2007). Defining What Digital Curators Do and What they Need to Know: The DigCCurr Project. *Proceedings of the 2007 ACM/IEEE Joint Conference on Digital Libraries (JCDL 2007)*, 49-50.

Lee, T., Zhongyuan, W., Wang, H., & Hwang, S. (2011). Web Scale Entity Resolution using Relational Evidence (Technical Report No. MSR-TR-2011-30). Microsoft Research. Retrieved from http://research.microsoft.com:8082/pubs/145839/paper.pdf

Li, B., Appling, D. S., Lee-Urban, S., & Riedl, M. O. (2012). Learning Sociocultural Knowledge via Crowdsourced Examples. *Proceedings of the 4th AAAI Workshop on Human Computation*.

Lin, D., & Pantel, P. (2001). Discovery of Inference Rules for Question Answering. *Natural Language Engineering*, *7*(4), 343-360.

Linde, C. (2001). Narrative and Social Tacit Knowledge. *Journal of Knowledge Management*, *5*, 160-170.

Liu, S. B. (2010). Trends in Distributed Curatorial Technology to Manage Data Deluge in a Networked World. *UPGRADE: The European Journal for the Informatics Professional*, *11*(4), 18-24.

Longuet-Higgins, H. C. (1973). Comments on the Lighthill Report and the Sutherland Reply. *Artificial Intelligence: A Paper Symposium*.

Lynch, J. G., Jr. & Srull, T. S. (1982). Memory and Attentional Factors in Consumer Choice: Concepts and Research Methods. *Journal of Consumer Research*, *9*(1), 18-37.

Malone, T. W., Laubacher, R., & Dellarocas, C. (2009, February). *Harnessing Crowds: Mapping the Genome of Collective Intelligence* (Research Report No. Working Paper 4732-09). Cambridge, MA: MIT Sloan School of Management.

Mataric, M. J. (1993). Designing Emergent Behaviors: From Local Interactions to Collective Intelligence. *Proceedings, From Animals to Animats, Second International Conference on Simulation of Adaptive Behavior*, 432-441.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1996, April 3). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Retrieved December 9, 2012, from http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

McCorduck, P. (2004). *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence* (2nd ed.). A K Peters/CRC Press.

Miller, G. A. (2003). The Cognitive Revolution: A Historical Perspective. *TRENDS in Cognitive Sciences*, *7*(3), 141-144.

Minsky, M. (1982). Why People Think Computers Can't. *AI Magazine*, *3*(4).

Munro, R., Bethard, S., Kuperman, V., Lai, V. T., Melnick, R., Potts, C., Schnoebelen, T., Tily, H. (2010). Crowdsourcing and Language Studies: the New Generation of Linguistic Data. *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (CSLDAMT '10)*, 122-130.

Negley, G. (1951). Cybernetics and Theories of Mind. *The Journal of Philosophy*, *48*(19), 574-582.

Nguyen, C. (2015). Algorithms of the Mind: What Machine Learning Teaches Us About Ourselves. Retrieved June 23, 2015, from https://medium.com/deep-learning-101/algorithms-of-the-mind-10eb13f61fc4

Nickerson, J. V., Zahner, D., Corter, J. E., Tversky, B., Yu, L., & Rho, Y. J. (2009). Matching Mechanisms to Situations Through the Wisdom of the Crowd. *Proceedings of the International Conference on Information Systems (ICIS 2009)*.

Nilsson, N. J. (1980). The Interplay Between Experimental and Theoretical Methods in Artificial Intelligence. *Cognition and Brain Theory*, *4*(1), 69-74.

Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation. *Long Range Planning*, *33*(1), 5-34.

O'Donovan, J. (2009). Capturing Trust in Social Web Applications. In J. Golbeck (Ed.), *Human-Computer Interaction Series: Computing with Social Trust* (pp. 213-257). Springer.

Parameswaran, M., & Whinston, A. B. (2007). Research Issues in Social Computing. *Journal of the Association for Information Systems*, *8*(6), 336-350.

Parameswaran, M., & Whinston, A. B. (2007). Social Computing: An Overview. *Communications of the Association for Information Systems*, *19*, 762-780.

Parker, L. E. (2008). Distributed Intelligence: Overview of the Field and its Application in Multi-Robot Systems. *Journal of Physical Agents*, *2*(1), 5-14.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45-77.

Pickering, A. (2011). *The Cybernetic Brain: Sketches of Another Future*. Chicago, IL: University of Chicago Press.

Pollock, J. (1990). Philosophy and Artificial Intelligence. *Philosophical Perspectives*, *4*, 461-498.

Rao, R. (2003). From Unstructured Data to Actionable Intelligence. *IT Professional*, *5*(6), 29-35.

Ravichandran, D., & Hovy, E. (2002). Learning Surface Text Patterns for a Question Answering System. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 41-47.

Reliability of Wikipedia. (n.d.). In *Wikipedia*. Retrieved June 26, 2015, from https://en.wikipedia.org/wiki/Reliability_of_Wikipedia

Rotman, D., Procita, K., Hansen, D., Parr, C. S., & Preece, J. (2012). Supporting Content Curation Communities: The Case of the Encyclopedia of Life. *Journal of the American Society for Information Science and Technology*, *63*(6), 1092-1107. http://dx.doi.org/10.1002/asi.22633

Sagan, C. (1986). *The Dragons of Eden: Speculations on the Evolution of Human Intelligence*. Ballantine Books.

Sarnovsky, J. (2006). Modern Rationality: A Cybernetic View. *Kybernetes*, *35*(10), 1745-1755.

Sarter, N. B. (2006). Multimodal Information Presentation: Design Guidance and Research Challenges. *International Journal of Industrial Ergonomics*, *36*, 439-445.

Sato, K. (1991). From AI to Cybernetics. *AI & Society*, *5*(2), 155-161. doi:10.1007/BF01891721

Schunn, C. D., Crowley, K., & Okada, T. (1998). The Growth of Multidisciplinarity in the Cognitive Science Society. *Cognitive Science*, *22*(1), 107-130.

Schwenk, C. R. (1988). The Cognitive Perspective on Strategic Decision Making. *Journal of Management Studies*, 25(1), 41-55.

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, *34*(1), 1-47.

Seigneur, J.-M. (2009). Social Trust of Virtual Identities. In J. Golbeck (Ed.), *Human-Computer Interaction Series: Computing with Social Trust* (pp. 73-118). Springer.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27*, 379-423, 623-656.

Shedroff, N. (1999). Information Interaction Design: A Unified Field Theory of Design. In R. Jacobson (Ed.), *Information Design* (p. 267). Cambridge, MA: The MIT Press.

Singh, M. D., & Kant, R. (2008). Knowledge Management Barriers: An Interpretive Structural Modeling Approach. *International Journal of Management Science and Engineering Management*, *3*(2), 141-150.

Soricut, R., & Brill, E. (2006). Automatic Question Answering Using the Web: Beyond the Factoid. *Journal of Information Retrieval*, *9*(2), 191-206.

Speier, C., Vessey, I., & Valacich, J. S. (2003). The Effects of Interruptions, Task Complexity, and Information Presentation on Computer-Supported Decision-Making Performance. *Decision Science*, *34*(4), 771-797.

Tarasewich, P., & McMullen, P. R. (2002). Swarm Intelligence: Power in Numbers. *Communications of the ACM*, *45*(8), 62-67.

Thagard, P. (2009). Why Cognitive Science Needs Philosophy and Vice Versa. *Topics in Cognitive Science*, *1*(2), 237-254.

Tractinsky, N., & Meyer, J. (1999). Chartjunk or Goldgraph? Effects of Presentation Objectives and Content Desirability on Information Presentation. *MIS Quarterly*, *23*(3), 397-420.

Vaishnavi, V., & Kuechler, B. (2004, January 20). Design Science Research in Information Systems. Retrieved March 31, 2013, from http://www.desrist.org/design-research-in-information-systems/

Valiant, L. G. (1984). A Theory of the Learnable. Communications of the ACM, 27(11), 1134-1142. http://dx.doi.org/10.1145/1968.1972

Vaughan-Nichols, S. J. (2012, December 17). Wikipedia Moving from MySQL to MariaDB. Retrieved from http://www.zdnet.com/wikipedia-moving-from-mysql-to-mariadb-7000008912/

Vicknair, C., Macias, M., Zhao, Z., Nan, X., Chen, Y., & Wilkins, D. (2010). A Comparison of a Graph Database and a Relational Database. *Proceedings of the 48th Annual Southeast Regional Conference*.

Vinyals, O. & Le, Q. V. (2015). A Neural Conversational Model. *Proceedings of the 31st International Conference on Machine Learning*.

Waltz, D. L. (1978). An English Language Question Answering System for a Large Relational Database. *Communications of the ACM*, *21*(7), 526-539.

Wang, F.-Y., Zeng, D., Carley, K. M., & Mao, W. (2007). Social Computing: From Social Informatics to Social Intelligence. *IEEE Intelligent Systems*, *22*(2), 79-83.

Wang, Y., Zhu, M., Qu, L., Spaniol, M., & Weikum, G. (2010). Timely YAGO: Harvesting, Querying, and Visualizing Temporal Knowledge from Wikipedia. *Proceedings of the 13th International Conference on Extending Database Technology (EDBT '10)*, 697-700.

Weglarz, G. (2004, September 1). Two Worlds of Data - Unstructured and Structured. *Information Management*. Retrieved from http://www.information-management.com/issues/20040901/1009161-1.html

Wenger, E. (2000). Communities of Practice and Social Learning Systems. *Organization*, *7*(2), 225-246.

Wiener, N. (1961). *Cybernetics: On Control and Communication in the Animal and the Machine* (2nd ed.). New York, NY: The M.I.T. Press.

Wikimedia Users. (n.d.). In *Wikimedia*. Retrieved June 25, 2015, from https://strategy.wikimedia.org/wiki/Wikimedia_users

Wolpert, D. H., & Tumer, K. (2008, February). *An Introduction to Collective Intelligence* (Technical Report No. NASA-ARC-IC-99-63). NASA Intelligent Systems Division.

Wu, H., Zubair, M., & Maly, K. (2006). Harvesting Social Knowledge from Folksonomies. *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia (HYPERTEXT '06)*, 111-114.

Yakel, E., Conway, P., Hedstrom, M., & Wallace, D. (2011). Digital Curation for Digital Natives. *Journal of Education for Library and Information Science*, *52*(1), 23-31.

APPENDIX A: A Brief History of Select Cognition Disciplines

Social Learning Systems find referent value in several distinct, yet related fields of study. Where the *Conceptual Framework* section of this paper focuses on the reference disciplines most directly relevant to the development of a working model of an SLS, this appendix walks readers through a much broader overview of the computer science cognition disciplines and their respective histories (albeit in an undeservedly concise manner, though of course readers can look elsewhere for more thorough treatment of these subjects).

i. Information Theory and Cybernetics

Though information theory may not be the most obvious candidate to make an appearance within an overview of cognition disciplines, it would be a great disservice to relay the story of cybernetics, however briefly, without it having been introduced. Many will recognize Claude Shannon as the father of information theory, due predominantly to his publication of "A Mathematical Theory of Communication" (Shannon, 1948) – the *A* in the title later being substituted with *The* in recognition of the work's prominence – though it was in fact two decades earlier that Ralph Hartley expressed the notion of information as a quantifiable phenomenon (Hartley, 1928) with an almost mechanistic operational capacity. This ability to view information quantitatively and to exercise formal logics upon it became the essential bedrock of information theory. The receiving of widespread scholarly attention for the idea, however, would have to wait until the publication of Shannon's mathematical theory, which helped to solidify the emergent field's place in the history books.

In Shannon’s work, greater emphasis was given toward the technical capacities of various aspects of information transmission. This included a detailed explication of what would later be known as *Shannon entropy*, which establishes the fundamental limits of the lossless encoding of information. Importantly, from an information theoretic perspective, these concepts apply not only to, for example, data being transferred over a wire, but to all manner of information exchanges. In recognition of this, Christian (2011) observed that encoding (particularly lossy encoding) is the essence of language; it provides a translation from pure meaning to that which is capable of being communicated.

Speaking of Shannon's theory, Boden (2006, p. 285) noted that “instead of finding complexity inside the stimulus... they [information theorists] found it outside. That is, a stimulus was no longer definable in isolation. Much as John Dewey and Ralph Perry had seen stimulus as a covertly purposive term, so the information theorists saw it as covertly probabilistic”.

The timing of information theory’s arrival is also important due to its concomitance with the emergence of another discipline principally concerned with the transmission of informational stimuli: that is, *cybernetics*. Cybernetics has been defined as “the science of communication in animals, men and machines” (George, 1979) and “the science of steersmanship” (Pickering, 2011, p. 3). Norbert Wiener, a prominent cybernetician who made great strides in advancing the field, noted that in such a view, “we deal with automata effectively coupled to the external world, not merely by their energy flow, their metabolism, but also by a flow of impressions, of incoming messages, and of the actions of outgoing messages” (Wiener, 1961, p. 132). From these views of cybernetics, we may say that, where the information theoretic mind is concerned principally with the stimulus itself, cybernetics holds in higher regard the conduit.

Negley, another cybernetician, argued against confining scientific inquiry to the realm of that which could be quantified. “A more comprehensive understanding of experimental procedure would indicate that scientific method might be defined as that method of observation and formulation which produces the most precise and systematic results in terms of understanding and control of the data which are the object of scrutiny by the method” (Negley, 1951). This view aligned well with the overarching cybernetic view of the time, which was largely opposed to the mechanistic treatment of the mind (Sato, 1991). *Meaning*, to most cyberneticians, was a “counterfeit” concept whose essential nature could be easily mistaken for its objective appearance – that is, the stimuli themselves (Dupuy, 2000, p. 9).

Importantly, though cybernetics and information theory arise from somewhat disjoint premises, both served as significant catalysts for the propulsion of more scientific views of information interchange, and both, too, were of general applicability. As such, they quickly became central to evolving studies relating to cognition and the essential nature of the mind.

ii. Artificial Intelligence

Where the 1940’s welcomed the mainstream introduction of information theory and cybernetics, the 1950’s heralded the introduction of scholarly pursuits in “artificial intelligence”, which is generally believed to have been formally established at a Dartmouth conference in 1956. The term itself, however, predates this event by at least a year, having appeared in the conference proposal in the fall of 1955 (McCarthy et al., 1996).

The overarching purpose of this field was to explore the essential nature of intelligent functions, with the hope of arriving at more practical explanations for these phenomena (Nilsson, 1980). According to Coiera (1996), artificial intelligence researchers “work both to extend their understanding of the ways in which intelligent systems can be constructed and to apply that knowledge in the real world”. Though this didn’t necessarily have to mean machines that were capable of thinking or sentient automatons, the appeal of the fantastical abounded, and artificial intelligence quickly acquired a passionate following.

This early optimism, it turned out, may have had several undesirable repercussions. Three decades following the field's formal inception, AI was still widely viewed as a kind of ad hoc discipline that lacked scientific rigor. Hall and Kibler (1985) observed that the area had failed to ever assemble a “commonly accepted statement of purpose or description of conventional research practices”. Cohen and Howe (1989) observed the lack of a standard practice of evaluation in AI research, as well, driven in part by a lack of “formal research methods, standard experiment designs, and analytic tools”.

Even in the mid-nineties, Baldwin and Yadav (1995) reiterated the need for rigor in research inquiries into AI, observing that the field suffered from substantial methodological concerns. Over time, these concerns had become compounded through AI's having become a reference discipline to other areas of scholarly inquiry, including the emerging field of Management Information Systems (Evaristo and Karahanna, 1997).

"The main thing wrong with much work in AI", Pollock (1990) speculated, "is that it has not been based upon sound theoretical foundations. Providing these foundations is a matter of doing philosophy, and AI theorists need to learn more philosophy". It appeared, in retrospect, that the promise of the practical rewards of AI may have inadvertently posed a setback to its development as a rigorous scientific discipline. But, while the field's theoretical underpinnings and methodologies may have changed and become gradually more refined over time – owed, in part, to increased academic scrutiny – the pursuit of AI's "Holy Grail" continued in the form of its *strong* artificial intelligence program, which "commits to, and pursues, the possibility of developing artefacts which have minds in the sense that we take ourselves to have minds" (Carter, 2007).

Interestingly, where from the very beginning information theorists seemed to applaud AI's efforts, as evidenced by Claude Shannon's own participation as an organizer of the 1956 Dartmouth conference, some cyberneticians did not share in the enthusiasm. As Bynum (2010) relates, "[Norbert] Wiener worried about the possibility that machines that learn and make decisions might generate significant ethical risks". Marvin Minsky, a pioneer in the field of artificial intelligence, however, did not feel that the intelligences of the artificial and of humans needed to intrinsically resemble one another (McCorduck, 2004, p. 126). For example, concepts such as emotions or desire, as we think of them, may be of little utility for a "thinking" machine. As Carl Sagan (1986) once put it, while "anatomy is not destiny... it is not irrelevant either". And the anatomy of the intelligent machine is widely left to the discretion of its creator.

iii. Neuroscience

In the 1960’s, shortly after the popular advent of artificial intelligence, the *neuroscience* movement began to take hold, garnering a large volume of scholarly interest (Brook and Mandik, 2007). “Neuroscience”, (Thagard, 2009) explains, “operates below the psychological level, concerning itself with neural networks. Understanding of neurons often draws also on molecular processes, for example, how genes produce proteins within cells enabling the operations of neurotransmitters such as dopamine and serotonin”.

In turn, the *neural networks* that Thagard referred to represent “a dynamic system consisting of simple processing units, often called 'neurons' or 'nodes,' and information passing links between these nodes often called 'interconnects' or 'synapses,' which can perform information-processing by responding to a set of input nodes containing information requiring processing” (Greenwood, 2007). This suggests that neural networks represent a form of connectionism (Dupuy, 2000, p. 6), crafted to emulate the essential neural “circuitry” of the brain, as well as brain-like processing as a whole (Carter, 2007).

The neuroscientist, then, is one who has set out to pierce the long-standing veil of mystery surrounding the brain and nervous system, or at least to make as much progress as possible in the name of those pursuits. New discoveries in neuroscience, then, provide an empirical basis for understanding many complex biological and mental processes, which can be of considerable significance to the overwhelming majority of cognition-based disciplines.

For example, inspired by progress in the field of neuroscience, artificial neural network models were developed consisting of "layers of simple computing nodes that operate as nonlinear summing devices" (Dayhoff and DeLeo, 2001). At a more abstract level, these models can be expressed as formal specifications (e.g., mathematical), and their relevance to artificial intelligence, cybernetics, and information theory (to name just a few disciplines) is plain to perceive, and representative of the quintessential interdisciplinarity of the neuroscience movement.

iv. Cognitive Science

Having recently been thrust into the realm of the more organic and empirical by way of the neuroscience movement in the 1960's, the 1970's witnessed a dedicated and substantial push towards greater abstraction. In 1973, the field of artificial intelligence suffered an early "winter", or a period of decreased general funding, following the publication of the *Lighthill Report* at the request of the British Science Research Council. The report, being highly critical of AI's progress, garnered a large volume of feedback from academicians active in the field. Among these commentators was H. Christopher Longuet-Higgins, who in his response to the report (Longuet-Higgins, 1973), laid out a list of those fields mostly like to be "enriched by artificial intelligence studies". These included mathematics, linguistics, psychology, and physiology, which he collectively referred to as the "cognitive sciences", coining the name of what would soon become a new field of its own.

The real "cognitive revolution", however, may have occurred even earlier than this. George Miller, often ranked among the founders of cognitive psychology, dated the cognitive revolution in psychology to the early 1950's, noting that it was in fact a “counter-revolution”, where “the first revolution occurred much earlier when a group of experimental psychologists, influenced by Pavlov and other physiologists, proposed to redefine psychology as the science of behavior” (Miller, 2003).

Even with the cognitive revolution underway as early as the 1950’s, however, it wasn’t until that time two decades later when cognition was formally established as a science in its own right (Brook and Mandik, 2007). As Thagard (2009) notes, “the organizational beginnings of cognitive science in the late 1970s, heralded by formation of the journal Cognitive Science and the Cognitive Science Society, explicitly looked for research that combined psychology and artificial intelligence”.

Earlier, Schunn et al. (1998) enumerated as contributors to the inception of cognitive science the following disciplines: anthropology, artificial intelligence, education, engineering, human-computer interaction, linguistics, medicine, neuroscience, philosophy, psychology, sociology, and others. That this list differs in scope from the one of Longuet-Higgins indicates not a lack of intradisciplinary coherence, but rather increased cross-disciplinary significance, and the evolution of the field as a whole, over time.

Throughout this evolution, the core of cognitive science has remained mostly unchanged. Thagard (2009) described the field as one that provides understanding “by giving account of the nature of key phenomena such as inference”. Similarly, Bechtel observed that “mechanisms in cognitive science... are proposed to explain cognitive activities such as memory retrieval or problem solving by performing operations on representations that carry information about objects, events, and circumstances currently or previously encountered” (Bechtel, 2009).

From these observations, it is easy to see how cognitive science has positioned itself amongst the other disciplines discussed to this point. Where neuroscience intrinsically lends its focus to the material brain and nervous system, cognitive science is concerned with more abstract processes and representations of mental phenomena; where artificial intelligence may seek to create intelligent artifacts, cognitive science wishes to demystify the nature of intellect; and where cybernetic thought is independent of meaning, cognitivist thought, inextricably, is bound to it (Dupuy, 2000).

APPENDIX B: Select Referential Areas of Study

Below, a bulleted list of related areas of study is provided. This list does not adhere to any formal organizational schema, but offers a good starting point for researchers and practitioners interested in delving deeper into investigating the conceptual domains surrounding Social Learning Systems.

- **Collective Intelligence**
  - Distributed / Social Cognition
    - Swarm / Social Intelligence
    - Collective Behavior
    - Promise Theory
  - Social Neuroscience
    - Neuroanthropology
    - NeuroCulture
  - Computational Sociology
    - Social Simulation
    - Artificial Society
  - Stigmergy

- **Artificial / Synthetic Intelligence**
  - Machine Learning
  - Automated Reasoning
  - Question Answering
  - Autonomous / Intelligent Agents
    - Agent-based Models
    - Multi-agent Systems
    - Self-organization / Spontaneous Order / Emergence
  - Artificial Immune Systems
  - Artificial Life

- **Systems Theory**
  - Dynamical Systems
    - Complex Systems
    - Complex Adaptive Systems
    - Dynamical Systems Theory
  - Social Complexity
  - Economic Systems
    - Complexity Economics
  - Cybernetics

- **Decision Theory**
    - Control Theory
    - Development Theory
    - Probability Theory
        - Generative Science
        - Chaos Theory
            - Bifurcation Theory
            - Catastrophe Theory
    - Cognitive Decision Making
        - Bounded Rationality
        - Cognitive Bias
        - Cognitive Distortion
        - Cognitive Dissonance
        - Decision Field Theory
        - Decision Engineering
    - (G)DSS / Expert Systems
    - Game Theory
        - Equilibrium
        - Neuroeconomics

- **Philosophy of Mind**
    - Logics
        - Classical Logics (e.g., Boolean)
        - Non-classical Logics (e.g., Fuzzy)
        - Order Theory
        - Domain Theory
    - Computational Theory of Mind
        - Computational Learning Theory
        - Computational Intelligence
        - Evolutionary Computation
    - Epistemology
        - Phenomenology
        - Neurophenomenology
        - Constructivist Epistemology
        - Simulated Reality
        - Positivist Epistemology
        - Universal Darwinism
        - Evolutionary Epistemology
        - Behaviorism
        - Cognitivism
        - Situational Awareness / Assessment
            - Situated Cognition
            - Situational Intelligence
        - Quantum Mind
        - Quantum Cognition
        - Embodied Cognition

        - Post-cognitivism
        - Enactivism
        - Learning Theory
        - Social Learning Theory
        - Explanation-based Learning
    - Information Theory
        - Philosophy of Information
    - Connectionism

- **Neuroscience**
    - Neuroinformatics
    - Neural Networks
    - Neuro I.S.

- **Data Analysis (Analytics)**
    - Data Mining
    - Data Cleaning
    - Business Intelligence
    - Predictive Analytics
    - Social Network Analysis

- **Social Computing**
    - Crowdsourcing
        - Collaborative Filtering
        - Recommender Systems
        - Crowd Funding
    - Online Social Networks
    - New Media
        - Virality
    - Collaborative Editing
        - Digital / Social Curation
    - Virtual Worlds

- **Collaboration Science**

APPENDIC C: Information Presentation and Decision-Making

"Information makes data meaningful for audiences because it requires the creation of relationships and patterns between data" (Shedroff, 1999).

Information presentation is a crucial consideration for any sound implementation of a Social Learning System, due to the direct influence of SLS's on individual decision-making and information acquisition processes. This section provides a brief review of some of the research that has been conducted relating to information presentation and its effect on decision-making and information acquisition.

When making a decision, individuals often rely on an extensive amount of external information to assist in the performance of their mental accounting tasks, or to make filtering among options a more meaningful endeavor. The availability, accessibility, and presentation of information, therefore, act as vital components to the science underlying an individual's choices. How information is organized can have direct implications on consumer choice (Bettman et al., 1998), and when that information is not properly designed, it can lead to inefficient information processing (Horn, 1999).

When referring to the manner in which information is designed, information format is a critical area of interest. Information format may refer to either the precise layout of information on a display, or alternative means of conveying information to an audience. From a decision-making standpoint, information format can affect both the individual's option selection strategy, as well as the overall amount of information consumed during the decision process (Johnson, 1984). Specifics relating to information format choice are often contingent upon the overall objective of the information display. Tractinsky and Meyer (1999) note that, "when presenting information, the objective may be either to facilitate efficient decision making… or to strengthen one's social status (in line with business communication practices and self-presentation theories)”.

As one consequence, specific information presentation objectives may influence the availability of certain information, as well as have an impact on how information is emphasized or made accessible to its audience. As another implication, these objectives will often influence the mode or channel of information display (i.e., whether information is visual, auditory, tactile, and so on in nature).

While all sensory experiences can provide input valuable to a decision-making process (Sarter, 2006), visual information presentations (whether through textual or graphical modes of expression) are perhaps the most commonly encountered. Horn (1999) suggests that “many ideas are best expressed with visual language, and others can only be expressed by visual language”. In a similar spirit, Speier et al. (2003) have stated that “presenting information in ways that enhances the use of perceptual processes… facilitates the acquisition and processing of complex information”.

There are, however, unique decision-making considerations that must be kept in mind for all sensory mediums, and sight is no exception. Among these considerations are several biological and cognitive factors. At a high level, biological considerations may consist of whether or not a particular mode of sensory experience is available to begin with, or if it is otherwise impaired by natural or environmental phenomena. As another example, others have noted that, due to the influence of presentation format and learning goals on information processing, the memory structure of stored information and the recall facilities of that information may themselves be altered to accommodate particular objectives (Biehal and Chakravarti, 1982). In addition, there are a great variety of biological factors that are much more narrowly focused in their nature.

In the case of visual sensory experiences, the gap between temporal and spatial resolution of the human visual system (where temporal resolution is considerably less than that of spatial) is noteworthy (de Bruijn and Spence, 2000), as it emphasizes the importance of efficient information conveyance in visual displays. This may offer some explanation as to why graphical representations of information have been shown to allow individuals to process information more quickly (that is, due to a lack of temporal depth), but not necessarily more accurately (Chau et al., 2000). This can affect the overall format of presented information (including text).

As one example of how spatial resolution has been manipulated to compensate for deficits in visual-temporal resolution, Cooke (2005), through an analysis of 40 years of convergent media, highlighted a trend toward increasingly more scannable information presentations, relying more heavily on visual display components and purposeful boundaries between contents. Another example can be found in the development of information presentation techniques that are designed to exploit the power of human spatial resolution, such as Rapid Serial Visual Presentation (RSVP), which presents text or graphics at a fixed focal point in rapid succession.

With regards to techniques such as these, “tests of implicit perception have shown that often more information about a briefly presented visual stimulus is available than can be reported by the observer” (de Bruijn and Spence, 2000). This alleged ability for information to “stick” in short bursts may well be what fuels the promise of techniques such as RSVP, but the same authors have noted that there are often downsides, as well.

For example, a phenomenon known as *attentional blink* has been described, for which "identification of one target may interfere with the identification of subsequent targets," at least over very short periods of time (de Bruijn and Spence, 2000).

Importantly, there are downsides to any mode of visual information presentation. At a general level, choices may suffer from attentional biases in which an individual is too consumed with some existing thought or detail to be more wholly aware of their situation. This, in turn, can have far-reaching implications on the decision process. One example of an attentional bias affecting visual sensory experiences relates to the novelty of a perceived option. As Lynch and Srull (1982) have indicated, "one's attention is captured by information that is novel or inconsistent with a prior expectation," which can result in greater recall relating to the novel concept or item later on, though at a potentially considerable cost to other immediate information. Another bias imposed more directly by presentation choices is that information presented in close proximity or in a similar style as other information (within a given context) is seen as being related, whereas information that is separated or distinct in its related stylings is viewed as unrelated or disjoint (Bateman et al., 2001). Additional concerns relating to information accessibility and misalignments between presentation objectives and those of the decision-maker have been noted previously in the academic literature.

With keeping these various considerations pertaining to information presentation mode, format, and so on in mind, we are able to develop a much greater appreciation of the relationship that exists between information presentation and decision-making. Throughout this discourse, however, we have mostly overlooked the more optimal case, where decision-making processes are enhanced via the availability and accessibility of information and the strategic or creative processes that have resulted in its eventual presentation format. This “optimal” case can be described in terms of achieving a *cognitive fit*, which occurs when the presentation format chosen allows the consumer to most effectively complete their task. This “facilitates decision making because the problem-solving processes used to act on the problem representation are similar to those needed to solve the problem” (Speier et al., 2003).

Clearly, this notion of cognitive fit is something of a moving target, having the ability to change in nature from one individual to the next. It is useful, however, for those in a position to facilitate information processing by way of information presentation to have an appreciation for the role of presentational aspects within the decision-making process, as we've discussed throughout this section.

APPENDIX D: A Unifying Philosophy of Information

While empiricism (discussed in the "Conceptual Framework" section of this text) may, in some light, serve as a common thread connecting the cognition-related disciplines, it stops short of providing a robust core of philosophical ideals, which may in turn arouse a number of cross-disciplinary incongruences, a handful of which have been identified previously in *Appendix A*. Sarnovsky (2006) suggested that progress in the related fields of cognition presupposes more "fundamental discoveries in logic", built upon an "immense reservoir of philosophy".

Brook and Mandik observed a movement involving the application of neuroscientific understanding to traditional philosophical questions. The central idea behind this movement was that some of these questions could only be answered by "a philosophically sophisticated grasp of... how the human brain processes information" (Brook and Mandik, 2007). Similarly, computationalism has been offered as a way of viewing the mind as an instantiation of "a particular formal system or collection of systems," where "mental operations are held to be computations" (Carter, 2007). Both of these approaches, however, seem to position the brain or mind themselves as controlling stakeholders in the new philosophy. While a new philosophy must certainly make adequate accommodations for the mind and its related operations, being confined to it would likely prove a devastating design flaw, as the principles of cognition and information sciences spread ever outward into new domains.

The Philosophy of Information (PI) is a relatively new contender that seems highly appropriate for accommodating this purpose. PI has been defined as "the philosophical field concerned with (a) the critical investigation of the conceptual nature and basic principles of information, including its dynamics, utilisation, and sciences, and (b) the elaboration and application of information-theoretic and computational methodologies to philosophical problems" (Floridi, 2002).

In PI, anything can be seen in the form of its informational content, which bodes well for both information theory, as well as neuroscientific and artificial intelligence-related views of the quantifiable and mechanistic mind. As Bynum (2010) observed, "a human is essentially a pattern of physical information, which endures over time, in spite of the constant exchange of molecules that occurs through biological metabolism". PI wishes to bring that physical information, as well as symbolic information, such as the mental representations of thought (Pollock, 1990), to the forefront of philosophical inquiry.

The significance of this view grows even greater when one considers that this "information" does not exist in stasis or in isolation. In fact, as George (1979) notes, we may even view thinking itself as "a process of manipulating symbolic representations of events, and the process of learning and adapting as a result of these manipulations". Information, then, becomes the universal currency of inquired things, whether they be sentient or otherwise, or even physical or intangible.

In adopting this perspective, "informational and computational concepts, methods, techniques, and theories... become powerful metaphors acting as 'hermeneutic devices' through which to interpret the world" (Floridi, 2002). The power of the hermeneutic devices of PI becomes apparent when challenged with difficult questions pertaining to cognition or other phenomena. One such example of this can be demonstrated through a PI-oriented explanation of creative thought, which has long been an uneasy terrain to navigate for several cognition-related disciplines. Artificial intelligence pioneer Marvin Minsky wrote that “we're so accustomed to the marvels of the unusual that we forget how little we know about the marvels of ordinary thinking. Perhaps our superstitions about creativity serve some other needs, such as supplying us with heroes with such special qualities that, somehow, our deficiencies seem more excusable” (Minsky, 1982). Several decades earlier, educational reformist John Dewey was likely to have shared this view, noting that “an individual can learn to think only in the sense of learning to employ more economically and effectively powers he already possesses” (Dewey, 1910).

From a perspective rooted in the Philosophy of Information, we can begin to offer a potential explanation in alignment with these two views: that creativity could be interpreted as a set of learned informational contents whose membership can be manipulated, in certain conjunctions with one another, to produce seemingly original outputs in alignment with the individual's current state of mind and available mental faculties. Here, we should note that by “state of mind”, we can mean either a common interpretation or a more rigorous view, such as Sagan’s (1986) observation of the human brain being capable of some 2 to the power of 10 trillion distinct *states* at any point in time. (The overwhelming magnitude of this number makes it highly unlikely that any two beings in existence – or, in fact, to have ever existed – are anything less than entirely unique in their mental constitution, and therefore their creative potential.) We might say of creative thought, then, that it is a perfect medley of reference, derivation, and juxtaposition of known information that creates information anew, and in potentially endless abundance.

In adopting the philosophy of information, even very daunting research questions can be more cleanly reduced to their essential nature, manipulated in an information theoretic way, and ultimately mapped back to their underlying material (or otherwise intangible) counterparts. This offers PI as a highly valuable candidate for incorporation in future works involving the design or employment of learning systems.

APPENDIX E: Tables

Table 1

Revised WordNet pointers used within the SLS demonstration

ID	Pointer	Description
1	Antonym	Related subject is opposite in meaning
2	Hypernym	Related subject is more general than this subject
3	Instance Hypernym	This is an instance of a related entity
4	Hyponym	Related entity is more specific than this
5	Instance Hyponym	Related entity can be an instance of this subject
6	Member/Part Holonym	Related entity is a whole that can contain this entity
7	Substance Holonym	Related entity is made from this entity
8	Member/Part Meronym	This entity is a whole that can contain the related entity
9	Substance Meronym	This entity is made with the related entity
10	Attribute	This noun's concept is described by the related entity
11	Derivationally Related Form	The related word is derived from this entity
12	Domain of Synset (Topic)	This entity belongs to the related domain
13	Member of Domain (Topic)	The related entity belongs to this domain
14	Domain of Synset (Region)	The related entity is regionally associated with this one
15	Member of Domain (Region)	This entity is regionally associated with the related entity
16	Domain of Synset (Usage)	The related entity classifies the usage of this entity
17	Member of Domain (Usage)	This entity classifies the usage of the related entity
18	Entailment	This entity is presupposed by the related entity
19	Cause	The related entity is a result of this entity
20	Also See	The related entity has a generic association with this entity
21	Verb/Adjective Group	This entity is similar in nature to the related entity
22	Verb Participle	This entity is derived from the related verb
23	Pertainym	The related entity is classified by or formed from this entity
24	Derived from Adjective	This entity is derived from the related adjective
25	Synonym	The related entity is synonymous with this entity

Table 2

Raw, tabular output from a prototypal SLS for the query "Albert Einstein"

Description	Albert Einstein (March 14, 1879 – April 18, 1955) was a German-born theoretical physicist. He developed the general theory of relativity, one of the two pillars of modern physics (alongside quantum mechanics). Einstein's work is also known for its influence on the philosophy of science. Einstein is best known in popular culture for his mass–energy equivalence formula $E = mc^2$ (which has been dubbed "the world's most famous equation"). He received the 1921 Nobel Prize in Physics for his

	"services to theoretical physics", in particular his discovery of the law of the photoelectric effect, a pivotal step in the evolution of quantum theory. Near the beginning of his career, Einstein thought that Newtonian mechanics was no longer enough to reconcile the laws of classical mechanics with the laws of the electromagnetic field. This led to the development of his special theory of relativity. He realized, however, that the principle of relativity could also be extended to gravitational fields, and with his subsequent theory of gravitation in 1916, he published a paper on general relativity. He continued to deal with problems of statistical mechanics and quantum theory, which led to his explanations of particle theory and the motion of molecules. He also investigated the thermal properties of light which laid the foundation of the photon theory of light. In 1917, Einstein applied the general theory of relativity to model the large-scale structure of the universe. He was visiting the United States when Adolf Hitler came to power in 1933 and, being Jewish, did not go back to Germany, where he had been a professor at the Berlin Academy of Sciences. He settled in the U.S., becoming an American citizen in 1940. On the eve of World War II, he endorsed a letter to President Franklin D. Roosevelt alerting him to the potential development of "extremely powerful bombs of a new type" and recommending that the U.S. begin similar research. This eventually led to what would become the Manhattan Project. Einstein supported defending the Allied forces, but largely denounced the idea of using the newly discovered nuclear fission as a weapon. Later, with the British philosopher Bertrand Russell, Einstein signed the Russell–Einstein Manifesto, which highlighted the danger of nuclear weapons. Einstein was affiliated with the Institute for Advanced Study in Princeton, New Jersey, until his death in 1955.Einstein published more than 300 scientific papers along with over 150 non-scientific works. On December 5, 2014, universities and archives announced the release of Einstein's papers, comprising more than 30,000 unique documents. Einstein's intellectual achievements and originality have made the word "Einstein" synonymous with "genius".
WordNet Part of Speech	Noun
WordNet Definition	Physicist born in Germany who formulated the special theory of relativity and the general theory of relativity. Einstein also proposed that light consists of discrete quantized bundles of energy (later called photons) (1879-1955)
WordNet Synonym	Einstein
Advisees	Ernst G. Straus Kurt Mendelssohn Leo Szilard Mahmoud Hessaby Muhammad Raziuddin Siddiqui Nathan Rosen
Advisors	Alfred Kleiner Heinrich Friedrich Weber
Alias	Einstein
Appears in Ranked Lists	Time 100: The Most Important People of the Century – 1999 Top 100 historical figures of Wikipedia – 9
Artwork on the Subject	Albert Einstein Memorial

Awards Won	Copley Medal - For his theory of relativity and his contributions to the quantum theory. – 1925 Franklin Medal - For his extensive work on relativity and the photo-electric effect - 1935 Gold Medal of the Royal Astronomical Society - 1926 Matteucci Medal - For fundamental contributions to the progress of science - 1921 Max Planck Medal - Max Planck - 1929 Nobel Prize in Physics - 1921 Nobel Prize - For his services to Theoretical Physics, and especially for his discovery of the law of the... – 1921
Book Editions Published	The World As I See It
Born	March 14, 1879 Ulm, Kingdom of Württemberg, German Empire
Cause of Death	Abdominal aortic aneurysm
Children	Eduard Einstein Hans Albert Einstein Lieserl Einstein
Citizenship	Kingdom of Württemberg (1879–1896), Stateless (1896–1901), Switzerland (1901–1955), Austria of the Austro-Hungarian Empire (1911–1912), Germany (1914–1933), United States (1940–1955)
Date of Birth	1879-03-14
Date of Death	1955-04-18
Departments	Humboldt University of Berlin Department of Physics School of Mathematics, Institute for Advanced Study
Died	April 18, 1955 (aged 76) Princeton, New Jersey, United States
Doctoral Advisor	Alfred Kleiner
Education	Bachelor of Science - 1901 - ETH Zurich - Mathematics and Physics – 1896 High school - 1896 - Aargau Cantonal School - 1895 PhD - 1905 - University of Zurich - Physics Primary school - 1895 - Luitpold Gymnasium
Employment History	Charles University in Prague ETH Zurich Institute for Advanced Study - 1933 - 1955 Leiden University Swiss Federal Institute of Intellectual Property - 1902 - Patent examiner - 1907 University of Zurich – 1909
Ethnicity	Ashkenazi Jews German American Germans Jewish people
Family	Einstein family
Fields	Physics, philosophy
Film Appearances	Atomic Power - Him/Herself Der ewige Jude - Archive Footage Elvis: Return to Tupelo - Archive Footage Hitler, a Career - Archive Footage How William Shatner Changed the World - Archive Footage Journey to Palomar, America's First Journey Into Space - Archive Footage On the Brink: Doomsday - Archive Footage Prophets of Science Fiction - Archive Footage Trinity and Beyond: The Atomic Bomb Movie - Archive Footage World Leaders on Peace and Democracy - Him/Herself

Follows Diet	Veganism Vegetarian food
Gender	Male
Hall of Fame Inductions	1990 - The Walhalla 2008 - New Jersey Hall of Fame
Height Meters	1.75
Influenced	Boris Podolsky Charles H. Bennett David Bohm Karl Popper Léon Brillouin Merce Cunningham Paco Ahlgren Rudolf Carnap Théophile de Donder Wolfgang Pauli
Influenced By	Baruch Spinoza Bernhard Riemann Ernst Mach George Bernard Shaw Hendrik Lorentz Hermann Minkowski Isaac Newton James Clerk Maxwell Karl Pearson Paul Valéry
Institutions	Swiss Patent Office (Bern), University of Zurich, Charles University in Prague, ETH Zurich, Caltech, Prussian Academy of Sciences, Kaiser Wilhelm Institute, University of Leiden, Institute for Advanced Study
Inventions	Einstein refrigerator
Known For	General relativity and special relativity, Photoelectric effect, $E=mc^2$, Theory of Brownian motion, Einstein field equations, Bose–Einstein statistics, Bose–Einstein condensate, Gravitational wave, Cosmological constant, Unified Field Theory, EPR paradox
Member Of	Royal Society Society of Friends of the Jewish National Library
Name	Albert Einstein
Namesakes	Albert Einstein Award Albert Einstein College of Medicine Albert Einstein High School Albert Einstein Medal Bose–Einstein condensate Einstein notation Einstein Observatory Einstein refrigerator Einstein tensor Tatung Einstein
Nationality	Austria-Hungary German Empire Germany Kingdom of Württemberg Switzerland United States of America Weimar Republic
Notable As	Academic

Notable Awards	Barnard Medal (1920), Nobel Prize in Physics (1921), Matteucci Medal (1921), ForMemRS (1921), Copley Medal (1925), Max Planck Medal (1929), Time Person of the Century (1999)
Notable For	Physicist
Official Website	http://www.einstein.biz/
Organization Board Memberships	Black Mountain College - Board of directors Society of Friends of the Jewish National Library - Vice Chair
Organizations Founded	Hebrew University of Jerusalem International Rescue Committee Pugwash Conferences on Science and World Affairs
Other Academic Advisors	Heinrich Friedrich Weber
Parents	Hermann Einstein Pauline Einstein
Peers	Constantin Carathéodory David Hilbert Edwin Hubble Erwin Schrödinger Kurt Gödel Niels Henrik David Bohr Werner Heisenberg
Place of Birth	Ulm
Place of Death	Princeton
Places Lived	Bern Germany Munich Princeton
Profession	Author Mathematician Philosopher Physicist Scientist Teacher Theoretical Physicist Writer
Quotations	"Creating a new theory is not like destroying an old barn and erecting a skyscraper in its place. It is rather like climbing a mountain, gaining new and wider views, discovering unexpected connections between our starting points and its rich environment. But the point from which we started out still exists and can be seen, although it appears smaller and forms a tiny part of our broad view gained by the mastery of the obstacles on our adventurous way up." "Great spirits have always encountered violent opposition from mediocre minds." "If men as individuals surrender to the call of their elementary instincts, avoiding pain and seeking satisfaction only for their own selves, the result for them all taken together must be a state of insecurity, of fear, and of promiscuous misery." "Imagination is more important than knowledge." "In the middle of difficulty lies opportunity."

	"Most of the fundamental ideas of science are essentially simple, and may, as a rule, be expressed in a language comprehensible to everyone." "Not everything that counts can be counted, and not everything that can be counted counts." "Technological progress is like an ax in the hands of a pathological criminal." "The ideals which have always shone before me and filled me with the joy of living are goodness, beauty, and truth." "This is what the painter, the poet, the speculative philosopher, and the natural scientists do, each in his own fashion."
Religion	Agnosticism Deism Judaism
Representations in Fiction	Albert Einstein
Residence	Germany, Italy, Switzerland, Austria (today: Czech Republic), Belgium, United States
Series Written or Contributed to	The Collected Papers of Albert Einstein
Sibling	Maja Einstein
Spouse	1903-01-06 - Bern - Mileva Marić - 1919-02-14 - Marriage 1919-06-02 - Elsa Einstein - 1936-12-20 – Marriage
Thesis	Eine neue Bestimmung der Moleküldimensionen (1905)
Weight kg	90.0
Works	A World Without Time: The Forgotten Legacy of Godel and Einstein Einstein Einstein in Love: A Scientific Romance Einstein, Picasso: Space, Time, and the Beauty That Causes Havoc Einstein: His Life and Universe Einstein's Clocks, Poincare's Maps: Empires of Time Einstein's Dreams Einstein's German World Einstein's Mistakes: The Human Failings of Genius Essential Einstein
Works Written	Albert Einstein, Hedwig und Max Born, Briefwechsel Albert Einstein/Mileva Marić - the Love Letters Bite-size Einstein Einstein's Annalen Papers Essential Einstein Ideas and Opinions La Mentalidad Militar Out of My Later Years Sidelights on Relativity The Born-Einstein Letters

APPENDIX F: Additional Screenshots

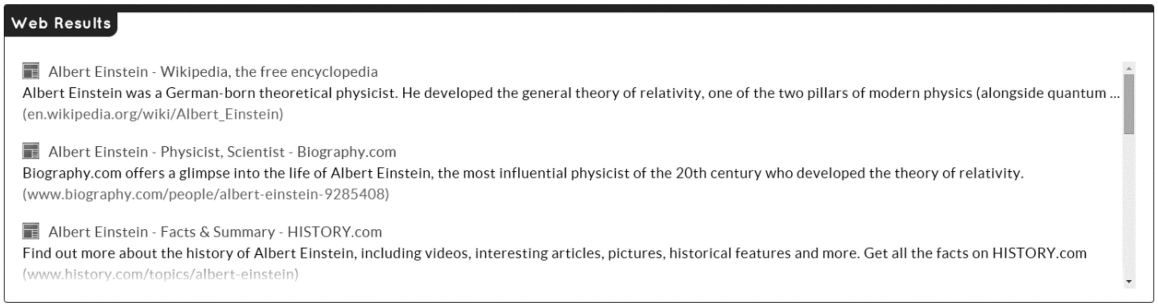


Figure 10. Web results panel generated by Calico.

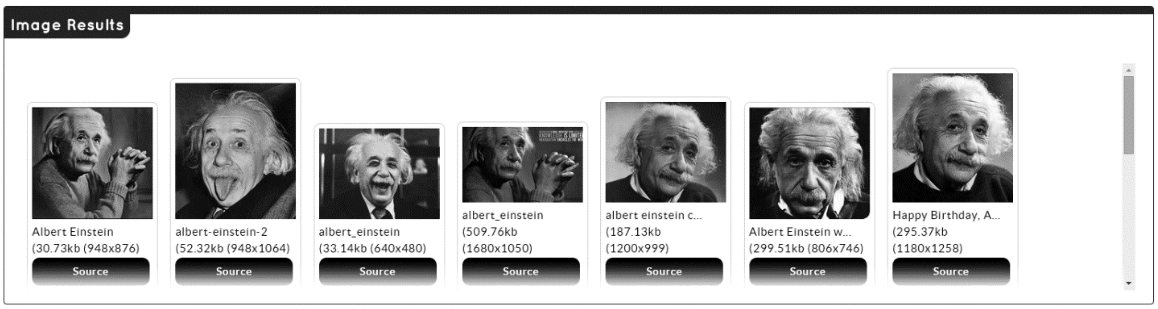


Figure 11. Image results panel generated by Calico.

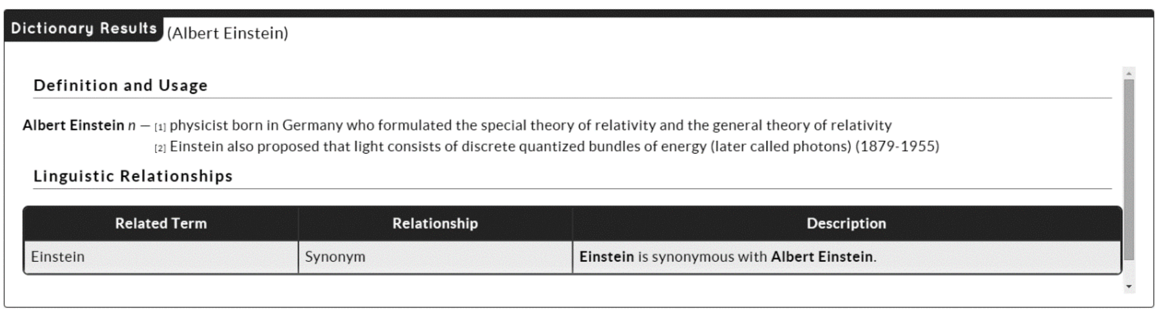


Figure 12. Dictionary panel generated by Calico.

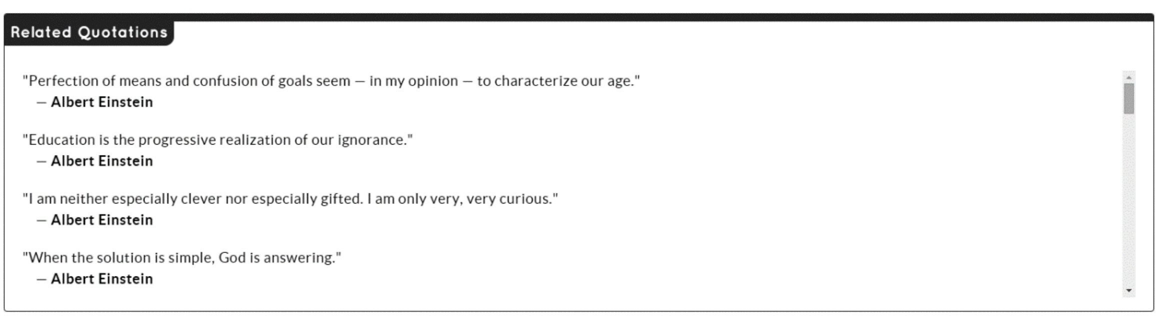


Figure 13. Quotations panel generated by Calico.

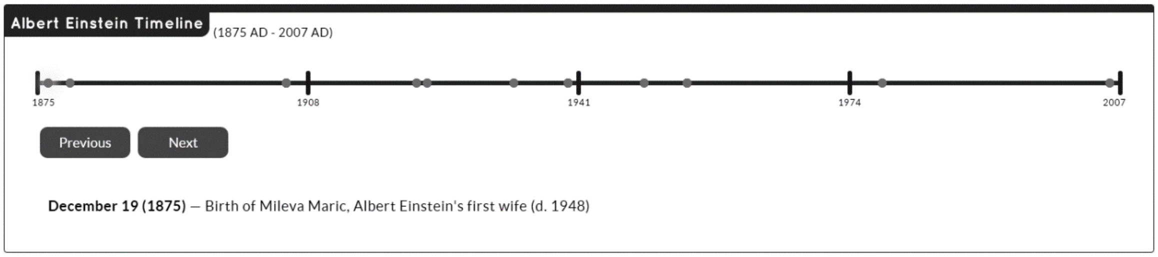


Figure 14. Timeline panel generated by Calico.

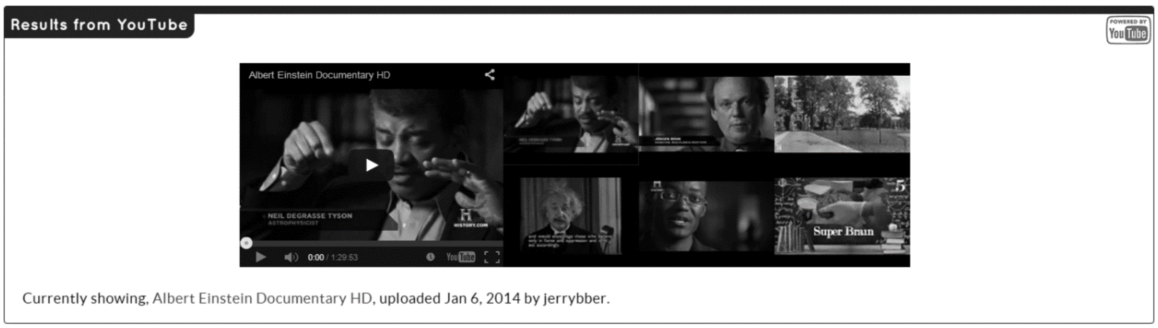
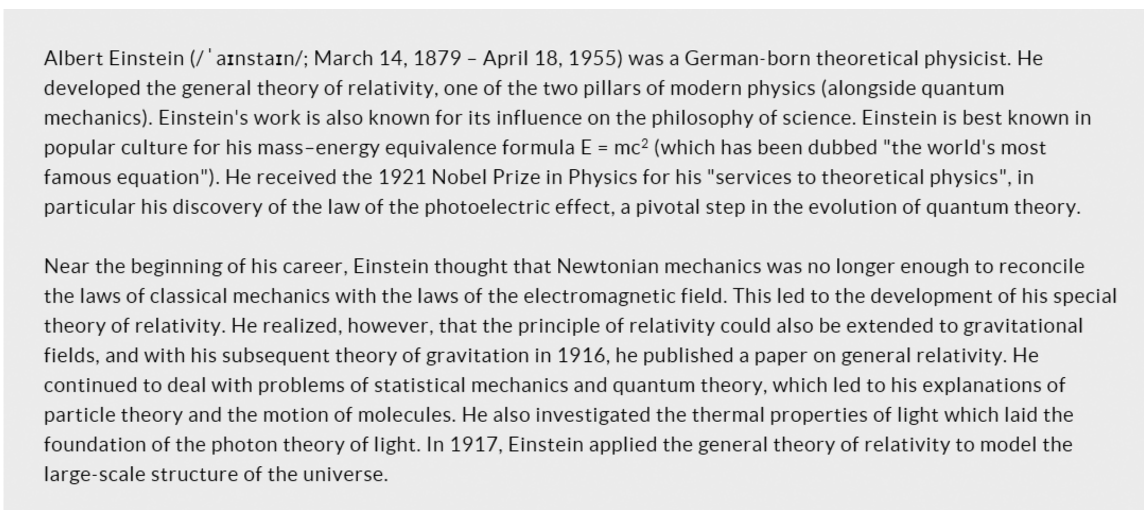


Figure 15. Video results panel generated by Calico.



Albert Einstein (/ˈaɪnstaɪn/; March 14, 1879 – April 18, 1955) was a German-born theoretical physicist. He developed the general theory of relativity, one of the two pillars of modern physics (alongside quantum mechanics). Einstein's work is also known for its influence on the philosophy of science. Einstein is best known in popular culture for his mass–energy equivalence formula $E = mc^2$ (which has been dubbed "the world's most famous equation"). He received the 1921 Nobel Prize in Physics for his "services to theoretical physics", in particular his discovery of the law of the photoelectric effect, a pivotal step in the evolution of quantum theory.

Near the beginning of his career, Einstein thought that Newtonian mechanics was no longer enough to reconcile the laws of classical mechanics with the laws of the electromagnetic field. This led to the development of his special theory of relativity. He realized, however, that the principle of relativity could also be extended to gravitational fields, and with his subsequent theory of gravitation in 1916, he published a paper on general relativity. He continued to deal with problems of statistical mechanics and quantum theory, which led to his explanations of particle theory and the motion of molecules. He also investigated the thermal properties of light which laid the foundation of the photon theory of light. In 1917, Einstein applied the general theory of relativity to model the large-scale structure of the universe.

Figure 16. Wikipedia entity description in Calico.

DOCTORAL ADVISOR	Alfred Kleiner
EDUCATION	• Bachelor of Science - 1901 - ETH Zurich - Mathematics and Physics - 1896 • High school - 1896 - Aargau Cantonal School - 1895 • PhD - 1905 - University of Zurich - Physics • Primary school - 1895 - Luitpold Gymnasium
EMPLOYMENT HISTORY	• Charles University in Prague • ETH Zurich • Institute for Advanced Study - 1933 - 1955 • Leiden University • Swiss Federal Institute of Intellectual Property - 1902 - Patent examiner - 1907 • University of Zurich - 1909
ETHNICITY	• Ashkenazi Jews • German American • Germans • Jewish people

Figure 17. Attributes view from Calico.

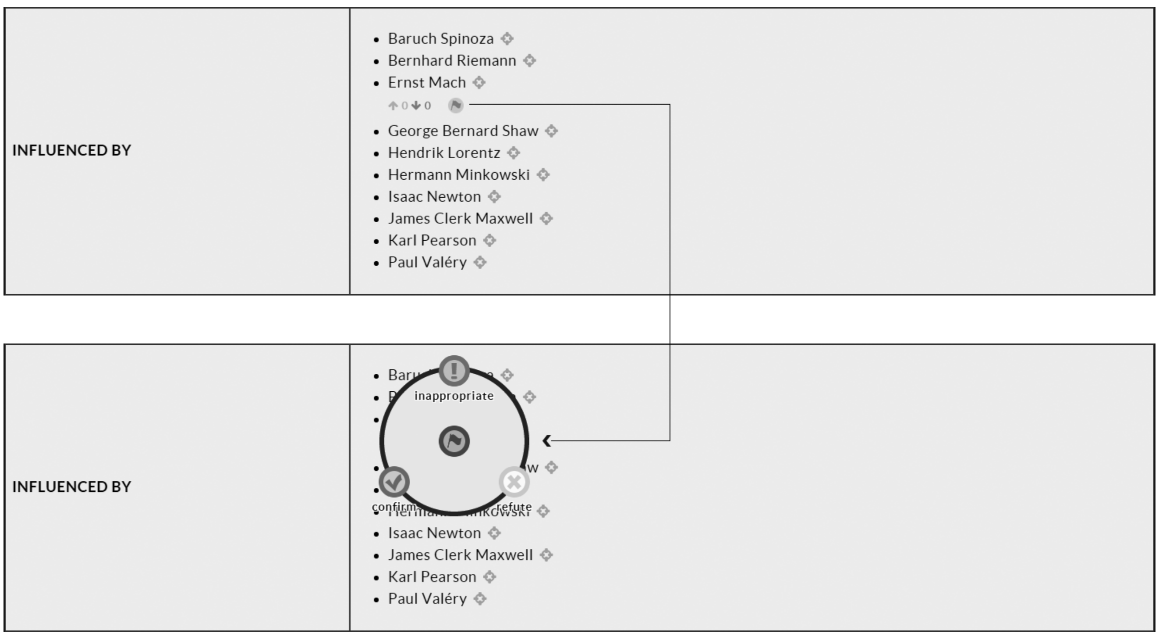


Figure 18. Voting and curatorial flagging in Calico.