Available Article

Cite as: Bergman, M. K. Potential Uses in Breadth. in *A Knowledge Representation Practionary: Guidelines Based on Charles Sanders Peirce* (ed. Bergman, M. K.) 319–341 (Springer International Publishing, 2018). doi:10.1007/978-3-319-98092-8_15

Official site: https://link.springer.com/book/10.1007/978-3-319-98092-8

Full-text: http://www.mkbergman.com/publications/akrp/chapter-15.pdf

Abstract: Four potential near-term applications are word sense disambiguation, relation extraction, reciprocal mapping, and extreme knowledge supervision. We next cover four logics and representations in automatic hypothesis generation, encapsulating KBpedia for deep learning, measuring classifier performance, and the thermodynamics of representation itself. Two of these, self-service business intelligence and semantic learning, have been on wish lists for years. The examples in this chapter show the benefits of organizing our knowledge structures using Peirce's universal categories and typologies.

Further, with its graph structures and inherent connectedness, we also have some exciting graph learning methods that we can apply to KBpedia and its knowledge bases.

15

POTENTIAL USES IN BREADTH

We begin this last *Part V* looking at *potential* applications. These knowledge management uses, made possible by following Peirce's guidelines, leverage KBpedia and domain extensions to it. I have assembled these examples to illustrate our intent in this *practionary* to attain what Peirce called "the third grade of clearness of apprehension":

> "It appears, then, that the rule for attaining the third grade of clearness of apprehension is as follows: Consider what effects, that might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object." (1878, CP 5.402)

It is, of course, impossible to conceive of all practical effects from a thing. However, in this chapter, and the one that follows, I try to share what I see as some important practical effects of applying Peirce's guidelines to knowledge representation. To my knowledge, few have implemented the ideas listed in this or the next chapter. The practical effects of these ideas are strong potentials with reasonable prospects for being realized. These ideas, collectively, help us begin to apprehend this 'third grade' of clear understanding.

I have selected these case examples both to highlight the diversity of potential uses and to showcase those with the highest likelihood of impact. Because what manifests in the future often 'surprises,' I am likely overlooking some impactful and practical effects of what may unfold in the future. Nonetheless, this method of selection does conform to what Peirce called the pragmatic maxim as a way to sift through the myriad of possible explanations for things to focus on those with the most economy and likelihood of bearing fruit.

I introduce each case with some context and a problem statement, then an introduction of concepts and existing building blocks that pertain to it, then to possible generalizations and potential practical effects were the case implemented. I do not exhaust the potential high-impact applications in these chapters. Recall we provided a long list of other possible uses of our approach in *Table 4-1* in *Chapter 4*. Consult that list for a fuller picture of potential applications. You will see, for example, that the case studies in this concluding *Part V* do not include applications such as ontologydriven applications (ODapps), concept alignment, entity and concept extraction, and semantic search, to name just a few of the important missing ones.

In this first concluding chapter, I briefly present about a dozen possibilities in *breadth* that introduce a variety of practical KR uses. These possibilities are more of an overview than the in-depth cases in the following chapter, but in their totality provide a good sense of potentials. We split these dozen possibilities into near-term potentials, logic and representations, and other more speculative potentials. The order of presentation is from the near-at-hand to the speculative. In the next *Chapter 16,* we present three practical applications, also pretty near at hand, in a considerable amount of *depth*. The combination of these two chapters of breadth and depth, the dimensions of Peirce's definition of information (*Chapter 2*), broadly captures the sense of practical and potential uses of our approach. *Chapter 17* concludes this last *Part V* of the book, re-capping Peirce's guidelines for knowledge representation.

NEAR-TERM POTENTIALS

We have already discussed how ontologies may drive bespoke applications and Web services. We have seen the importance of organizing attributes and mapping to them for instance characterization and intensionality. Four further potentials are also near at hand in word sense disambiguation, relations extraction, reciprocal mapping, and extreme knowledge supervision. These potential applications all are examples of leveraging the rich structure of KBpedia and its extensions.

Word Sense Disambiguation

Word sense disambiguation is picking the correct meaning for a word where it has multiple meanings.* Vocabularies grow by either minting new words or giving new meanings, also called *senses*, to existing words. Multiple senses for common words is a historical linguistic result of the bifurcated chaining of new word senses for new uses based on adjacent metaphors.[1] This mode of how new word senses get coined conforms to the least 'cognitive cost' for generating, interpreting, and learning them.[2] Some of these senses, such as *game* for *hunted fowl* or *game* for an *amusing pastime*, may have diverged long ago with a broad span of meaning.

The traditional approach to word sense disambiguation (WSD) uses dictionaries to look up the various senses of a word. Lesk is a leading method, wherein we search the various word senses in a dictionary based on the neighboring text for the search term. The Lesk algorithm calculates the overlap of the sense definition of a word and the contextual definitions of the terms that surround its use, with variants allowing us to control the sliding window or other parameters.[3] The limitation of the Lesk approach is that it depends on the wording of the definitions. We may also base word embeddings on other factors, including structure and other features.[4]

* WSD is also closely related to named-entity recognition or named-entity disambiguation. The dictionary basis shifts from word senses to entity characterizations (attributes), but much else in approach is similar.

Unsupervised learning surfaces other rules and insights useful to WSD. Nearly a quarter-century ago Yarowsky showed a strong tendency for one word sense per discourse and collocation.[5] Choosing the most frequent sense for a multi-sense term is one of the best performing heuristics.[6] On a more abstract front, Sun has shown a framework that regularizes the structure of feature-rich corpora, which can derive training models that can converge rapidly and reduce generalization risk.[7]

Methods for word sense disambiguation may also learn from large knowledge sources, with Banko and Brill one of the first.[8] One of their findings was that the larger the number of annotations for term entries, the better is the resulting accuracy. More recently, Ponzetto and Navigli have demonstrated that knowledge harvested from Wikipedia can be efficiently used to improve the performance of a WSD system.[9] Adding Wikipedia links to baseline approaches can further enhance disambiguation performance.[10] Still, WSD for state-of-the-art systems has 2% to 5% error, not including inter-annotator differences. These performance figures are also for very limited domains with corpora and training sets known in advance. Word sense disambiguation applied to new domains needs to overcome what is known as the knowledge acquisition bottleneck, which is the cost of finding, structuring or annotating knowledge for WSD and other natural language processing applications. Many difficulties occur in acquiring tagged senses for WSD.

The potential of KBpedia and how it is structured to improve the WSD picture is profound. First, we have an instance-rich knowledge structure. Not only does that structure bring direct benefits, but the hundreds to thousands of instances per type also provide a rich content base for various word- and sub-graph embedding models. Second, the KBpedia structure is coherent. Third, we base KBpedia on Peircean ideals of knowledge representation. Its features are mostly lexically based (relations, attributes, senses, and meanings), which means that abstraction layers through the use of neural nets have a higher prospect for being interpretable (and coherent). Fourth, because of the degree of semantic relatedness in the structure, chances are greater that neighbor-based methods to WSD will perform better than alternatives. Fifth, the KBpedia features, as *Appendix C* describes, are a richer base for structure regularization methods than what Sun has analyzed.[7]

So, what we see with a KBpedia-based approach to WSD is one that combines all of the best methods in a single package. Its contextual understandings can extend to entity recognition and disambiguation, as well as for concepts and relations. KBpedia's graph structure, with its emphasis on trichotomies and typologies, should also promise better performance because of its comparative simplicity and cleanliness. We have strong dictionary and synonym bases, combined with a coherent and robust graph structure with millions of instances with content, which is expandable for new domains, and testable with the potential for continuous improvement.

Relation Extraction

In the context of relation extraction, most define 'relation' as a form of connection between two objects. The objective of relation extraction, then, is to identify and extract this relation. In contrast, we have seen in the context of Peirce that a general relation may specialize into one of three forms: attributes, external relations, or representations. Our Peircean approach also gives us better tools to identify and extract general relations, and then to organize and reason over them.

Relation extraction attempts to correctly identify and extract what is essentially an RDF triple of *subject-predicate-object*. Sometimes the subject placeholder is blank or unknown; sometimes the object placeholder is blank or unknown. (Theoretically, we could also treat the predicate slot as a blank.) Because of these structural aspects to a relation, the earliest forms of extraction put forward by Hearst in 1992 used many heuristics applied to lexico-syntactic patterns.[11] These techniques are surprisingly effective for many relation patterns; many systems still use them. The kernel method builds on this approach by looking for patterns within generalized tuples. Supervised approaches can also work quite well since we can pose the problem as one of binary classification. Relation extraction was also one of the first applications of the use of knowledge bases to inform labeled examples, what we now call distant supervision,[12] which remains one of the better-performing methods. More recently, joint inference on both entities and the relations looks to improve extraction efficiency further.[13]

Relation extraction has some unique uses within NLP methods. First, of course, it is the method for extracting relations (though, as mentioned, this has not yet been distinguished from attributes and representations). Second, we may find patterns to help narrow the identification of new concepts or entities by analogy to existing complete patterns. In the most effective sense, we should be able to narrow the applicable types for the new concepts or entities as well, but that is little applied. The potential exists to improve significantly our ability to identify previously unseen entities, not already in our dictionaries or gazetteers. Third, because of its patterned nature, we also value relation extraction as a technique used in data mining and question answering.[14] Last, the potentials for relation extraction are even more vast, which I get to in a moment.

The TextRunner and then KnowItAll and ReVerb efforts from the Etzioni lab at the University of Washington, and more recently the Nell project from Carnegie Mellon University, have been mining Web sources to discover relations and their associated entities. These efforts use open information extraction as a technique for knowledge base population. These approaches are useful, for example, to identify new entity members for specific types, sometimes called 'slot filling,' with millions of candidates identified. Another application is to disambiguate entities based on context. Besides these university efforts, commercial entities have been doing the same. Still, relation extraction is a comparatively inaccurate NLP task due to the variability in the triples structure in language and the immense number of potential entities.

KBpedia can improve all aspects of extracting, identifying, reconciling, and organizing the three aspects of relations, which also should lead to new capabilities in ontology learning and better capabilities in question answering and data mining. Inspections of the object slot may also aid in error detection of values and other possible misassignments. We can realize these potentials due to the better characterization and structure of KBpedia.

If we someday want to create ontologies from raw input text, the dream of ontology learning, we will require broad and accurate characterizations of relations to decompose the meaning of text structure. We have already mentioned the fine-grained structure of relations in KBpedia. The three segments of attributes, external relations, and representations, organized by types, provide better structure for evaluating relations. An initial task is to inspect and map relations from VerbNet and the open IE projects. The Nell project also provides domains and ranges, which should be helpful to relate types to specific predicates. These characterizations, in turn, would enable better mapping and inference of entity types to predicates and other patterns. The feedback from this process would undoubtedly surface improvements to KBpedia, which would feedback into better extractions anew. Computerized machine reading or natural language understanding will need these capabilities. The area of relations extraction should be a fruitful research focus for many years to come.[15]

Reciprocal Mapping

The standard method of mapping is to relate new concepts and entities in an external knowledge 'source' (B) to the master or governing one in a 'target' resource (A). The use case typically uses the target resource as a reference for external sources, possibly for data federation or integration. The mapping statements take the form of A:B or B:A. However, the external source may also be a valuable contributor to new concepts or entities for the target resource. In this use case, our interest is adding more A' to A, rather than simply mapping statements. We call this use case 'reciprocal mapping,' a topic in *Chapter 13*. Reciprocal mapping is not warranted in all cases, and only best applies when we encounter a quite complete external source, as is the case of Wikipedia contributing to KBpedia.[16] It is also a particularly useful technique where one wants to augment an existing knowledge graph, perhaps in adding domain extensions to a starting basis in KBpedia.

First, let's assume that we have already mapped the matching concepts between B → A and B itself is a rich external source.* What we want to do is to use this linkage to propose a series of *new* sub-classes that we could add to A (KBpedia in our example case) based on the sub-categories that exist in B for each of these mappings. The challenge we face by proceeding in this way is that our procedure potentially creates tens of thousands of new candidates. Because the B category structure has an entirely different purpose than the KBpedia knowledge graph (A, in this case), and because B's creation rules are completely different from those of A (KBpedia), many candidates are inconsistent or incoherent to include. A cursory inspection shows that we should drop most of the candidate categories. It is not tenable to review hundreds of thousands of new candidates manually, as is the case when B is the size of Wikipedia; we need an automatic way to rank potential candidates.

Several factors differ for reverse (reciprocal) mapping from our standard B → A mapping case. First, we need to find missing clusters or new concepts or types in B

* Any sufficiently complete or robust external ontology closely related to the current domain needs may fulfill this role.

that fit, but are missing, in A. Second, we need to ensure the scope and boundaries of concepts or types in B are roughly equivalent to those in A. We may expend considerable effort to clean the source B type and category structure prior to the reciprocal mapping. Third, we also need to capture structural differences in the source knowledge graph (B). Possible category matches fall into three kinds: 1) *leaf* categories, which represent child extensions to existing KBpedia (A) terminal nodes; 2) *near-leaf* categories, which also are extensions to existing KBpedia terminal nodes, but which also are parents to additional child structure in the source; and 3) *core* categories, which tie into intermediate nodes in KBpedia that are not terminal nodes. By segregating these structural differences, we can train more precise placement learners.

We automate this process with an SVM classifier trained over graph-based embedding vectors generated using the DeepWalk method.[17] DeepWalk learns the subcategory patterns that exist in the B category structure in an unsupervised manner. The result is to create graph embedding vectors for each candidate node. Our initial B → A maps enable us to create training sets with thousands of pre-classified sub-categories quickly. We split 75% of the training set for training, and 25% for cross-validation. We also employ some hyperparameter optimization techniques to converge to the best learner configuration. Once we complete these three steps, we classify all of the proposed sub-categories and create a list of potential `subClassOf` candidates to add into KBpedia, which we then filter by relevance score and vet manually.

The reference 'gold' standards in the scored training sets (see *Chapter 14*) provide the basis for computing all of these statistics. We score the training sets as to whether a given mapping is true or false (correct or not). (False mappings should be purposefully introduced.) Then, when we parse the test candidates against the training set, we note whether the learner result is either positive or negative (indicated as correct or indicated as not correct). When we match the test to the training set, we thus get one of four possible scores: true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Those four simple scoring categories are sufficient to calculate any of the statistical measures, as we discussed in *Chapter 14*.

We capture the reciprocal mapping process using a repeatable pipeline with the reporting of these various statistical measures, enabling rapid refinements in parameters and methods to achieve the best-performing model. Once appropriate candidate categories are generated using this optimized model, we then manually inspect results and make final selections. We then run these selections against the logic and coherency tests for the now-modified graph and keep or modify or drop the final candidate mappings depending on how they meet the criteria. Our experience suggests this semi-automated process may take as little as 5% of the time it would typically take to conduct this process by comparable manual means.

So, machine learning methods may reduce the effort required to add new concepts or structure by 95% or more. Machine learning techniques can filter potential candidates automatically to reduce greatly the time a human reviewer has to spend to make final decisions about additions to the knowledge graph. A reusable pipeline leads to fast methods for testing and optimizing parameters used in the machine learning methods. We can systematically tune and rapidly vet this pipeline.

Extreme Knowledge Supervision

Recall from *Chapter 4* that *knowledge supervision* is the purposeful use and structuring of knowledge sources and graphs to provide features and training sets for *KBAI*. *Distant supervision* uses the same sources, though employed as is and not purposefully staged. In knowledge supervision, we design and prep the knowledge base so that its structure enables query selection of labeled positive (and, with repeatable techniques, negative) training sets for supervised machine learning. This pre-staging of the knowledge sources eliminates 80% of the effort or more required for most supervised learning tasks. We also showed a virtuous circle of interaction between properly designed knowledge bases and a knowledge graph such that we can add new assertions and facts to the knowledge base and improve its quality by a higher ratio of true positives (see *Figure 4-2*).

When repeatedly and purposefully carried out through many cycles, we can call this *extreme* knowledge supervision. In the case of KBpedia, remember, we already have important structural splits between concepts, entities, events, attributes, external relations, and representations, all organized according to the triadic *universal categories* of Charles Peirce, and further sub-typed by scores of modular *typologies*. Theoretically, we may use the intersection of any of these dimensions to create and train supervised learners. Further, because of this richness of structure, we also can develop better language parsers (see *Chapter 16*) and reasoners (see next) to apply to our tasks. Also, combinations of these features through inference over category structures is a patented way[18] that brings significant efficiencies.[19] Here is the breadth of tasks to which we may apply extreme knowledge supervision:

- Entity identification (recognition) and extraction;
- Attribute identification and extraction ('slot filling');
- Relation identification and extraction;
- Event identification and extraction;
- Entity classifiers;
- Phrase (n-gram) identification;
- Entity linkers;
- Mappers;
- Topic clusterers;
- Topic classifiers;
- Disambiguators;
- Duplicates removal;
- Semantic relatedness;
- Inference and reasoning;
- Sub-graph extraction;
- Ontology matchers;
- Ontology mappers;
- Sentiment analysis;
- Question answering;
- Recommendation systems;

- Language translation;
- Multi-language versions;
- Artificial writing; and
- Ongoing knowledge base improvements and extensions.

I have listed these areas in rough order from the simpler to the more complex analyses. Distant supervision efforts have concentrated on information extraction, the first items on the list. However, all are amenable to knowledge supervision with ML.

A vetted knowledge graph with millions of supporting instances also provides some graph-level benefits. The first area is in 'deep graphs.'[20] The basic idea behind 'deep graphs' is to segregate graph nodes and edges into *types*, which form supernodes and superedges, respectively. In our terminology, 'deep graph' node types are akin to types of similar *attributes,* and edge types are akin to types of *relations*. The 'deep graph' algorithm can partition these grouped types into lattices, which can be intersected (combinations of nodes and edges) into representing deeper graph structures embedded in the initial graph. We can use these deeper graph structures as new features for machine learning or other applications. A second area, important to data interoperability, is in 'symbol grounding'[21] (also see next chapter). The usefulness of symbol grounding resides in associating symbol tokens as understood by the computer with actual language meanings. Besides interoperability, such groundings are crucial to natural language understanding.

The idea of large knowledge bases providing enabling technology for knowledge sharing goes back at least 30 years.[22] We are still in the early phases of such iterative refinements of KBpedia. As this process continues, expect to see faster and more accurate learners, the incorporation of still-additional knowledge sources and datasets, and more sophisticated combinations of features and methods for extreme knowledge supervision. Song and Roth provide an excellent current survey with hundreds of references for how machine learning based on using world knowledge may create such potentials.[23]

LOGIC AND REPRESENTATION

The previous section begins to scratch the surface for how KBpedia, as structured using the guidelines of Peirce, may improve many knowledge-based tasks, especially in the areas of natural language processing (NLP). I would now like to move beyond this traditional baseline and address more fundamental questions of logic, reasoning, and representation. These kinds of fundamental questions can take the use and contributions of knowledge-based systems to new levels. The four initial topics we cover in this section include automatic hypothesis generation, encapsulating KBpedia for deep learning, measuring classifier performance, and the thermodynamics of representation itself.

I do not touch on all of the logical potentials in this section. For example, the use of fuzzy logic, or intensional logic, or methods of inductive reasoning provide enormous potential. Areas in non-classical logics such as three-valued logics[24] or triadic logic[25] also deserve attention given their relationship to Peirce's universal categories. These are worthy topics for future attention.

Automatic Hypothesis Generation

One of Peirce's signal contributions was to bring the importance of abductive reasoning to the fore in matters of epistemology. We discussed the now three classical logical methods of deduction, induction, and abduction in *Chapter 8*. *Deduction*, the most widely employed method in the semantic Web and knowledge graphs, evaluates correct placement by traceable logic chains, most of a hierarchical nature. *Induction*, little used but with great promise for knowledge graphs, can look to shared or common features to make probable assertions. *Abduction*, which Peirce brought to the fore, is the logic of new knowledge and scientific discovery. It is rarely used and not well understood, some due to Peirce's own changing views.

What Peirce early called abduction he later acknowledged was, in fact, induction. Peirce's confusion — and how he eventually worked out the issue — is instructive. What Peirce initially called abduction is what we now call inference to the best explanation (IBE). The basic idea is given a particular outcome, what is the most likely path through the knowledge graph that leads to that outcome? It is a form of backward chaining, where all parts of the syllogism are known, and therefore is a true inferential method. Still, many combinations are possible, and reasoning backward across available choices can soon become computationally intractable. Since in abductive reasoning we are ultimately seeking the explanation to a question or phenomenon, this kind of IBE reasoning is quite valuable for knowledge graphs in general[26] and has applicability to instance characterizations in the ABox as well.[27]

Still, this view of abductive reasoning is but a part of what Peirce intended in his mature formulation. Peirce was seeking no less than an understanding of how the scientific method (purposeful inquiry) worked and its logic, in a broad sense. His characterizations redound with expressions of 'surprising facts,' 'flashes,' 'guesses,' 'instinct,' and 'new knowledge.' Dewey, a fellow pragmatist, saw similar things, but particularly looked toward abductive reasoning also as a way to explicate learning.[28] Peirce well understood the combinatorial problem and sought to understand how we winnow through the myriad of options, recognizing the factors of economy, effort, the likelihood of producing results, and all of those things we now understand as 'pragmatic.' Peirce understood there was a transitional space between perception and hypothesis that held the key to this unique logic. Flach, throughout his many writings, has noted the importance of abductive and inductive logic to the development of scientific knowledge, and also usefully split Peirce's ideas of abduction into explanatory and confirmatory reasoning.29The nut to crack around abduction resides in explanatory reasoning. Flach has attempted to refine Peirce's conception of explanatory reasoning into a form amenable to logical analysis.[30]

Prying open the heart of the logic of science is an exciting prospect. Kapitan made a powerful argument for why IBE was not the nub of abductive reasoning, and suggested heuristic aids, while not inferences, could still be used to discover new knowledge, often based on analogy.[31] Kapitan also compiled eight reasons from Peirce for what we should seek in a candidate hypothesis to explain an observation or surprising fact:

1. The cost (in time, money, and effort) of testing the hypothesis (1901, CP 6.533; 1901, CP 7.230);
2. The intrinsic value of the hypothesis regarding its 'naturalness' and 'likelihood' (1901, CP 7.223);
3. The fact that the hypothesis can be readily broken down into and elements and studies (1901, MS 692:33);
4. The simplicity of the hypothesis (*i.e.*, it is more readily apprehended, more facile, more natural or instinctive) (1902, MS L75:286; 1901, CP 6.532; 1908, CP 6.477);
5. The breadth of the hypothesis or the scope of its predictions (1902, MS L75:241, 457:37);
6. The ease with which we may falsify the hypothesis (1902, MS L75:285);
7. The testability of the hypothesis using severe tests based on 'incredible predictions'; and
8. The analogy of the hypothesis with familiar knowledge (1901, MS 873:16).

These guidelines feel incomplete. As part of his treatment of logic within the universal categories, Peirce held abductive reasoning as irreducible from the other two forms of logic, deductive (2ns) and inductive (3ns). We are still missing the essence of what makes abductive reasoning different. If we can truly get at the essence of the scientific method and purposeful inquiry, we will have unlocked a tremendously powerful door to new knowledge and discovery.

Kapitan held that missing piece was the creative, what it is that underlies knowledge.[32] He did not see this as an inferential step, but as one 'suggested' by the facts, by a general cognition. Kapitan likened the transition from the perception that leads to the idea as arising by analogy, from the unconscious. Selected quotes by Peirce support parts of this interpretation.

More recently, Tschaepe questioned some of this interpretation, choosing to focus more on 'guessing.'[33] Successful guessing is both piecemeal and done in an orderly fashion, guided by ethics and aesthetics, situated to logic as Peirce did. Tschaepe notes that a more metaphorical kind of logic is in play, and is indeed playful ('musement' in Peirce's term). Some scholars see it as likely based on the detection of patterns. Yes, the process is logical in a broad sense but is also a rapid surfacing and evaluation of candidate explanations arising from patterned similarities and metaphors. This critical stage between perception and hypothesis evaluation is a multi-factorial, synthetic, broad contrast of iconic options rapidly screened for pragmatic likelihood. The methods of this critical phase in abduction appear more oriented to pattern matching than inference, which, in any case, appears weak. Once a potential hypothesis is chosen for some level of evaluation, it becomes indexical.

KBpedia, or its derivatives, has the raw grist to begin feeding tests of these broad factors. In the near- to intermediate-term, backward chaining and IBE look quite tractable within the KBpedia structure. Longer-term, however, getting at the true 'guessing' game involved with abductive reasoning — unique and broad — is where, I think, some surprisingly useful payoffs may result where KBpedia may contribute.

Encapsulating KBpedia for Deep Learning

Geoffrey Hinton is a founder of deep learning. He and his team at the University of Toronto helped promote the idea of backpropagation as a way to send weights to adjust supervised labels to unsupervised layers in a neural network, with the increasingly propagated layers leading to the term of deep learning. Deep learning is exceptionally effective for image and pattern recognition tasks, less so for natural language. Unfortunately, the representations at all layers of deep learning are opaque, meaning we can glean no meaning from the information at a given layer. This 'black box' aspect is the weakness of deep learning. The concern, of course, with methods that lack explanation is that it is hard to know how to make further improvements. Inexplicable methods always seem to top out at some limit of performance.

Hinton likely understands these limits better than anyone. Well before deep learning became such a buzz phrase, Hinton and his team in 2011 were experimenting with how to package features together to act as a unit during the deep learning process.[34] Hinton's group has been more focused on image representations than text. Still, this paper was the first mention of defining these feature packages as 'capsules.' Hinton has continued to work on this 'capsule' concept and has come to understand that clean features about single entities are the best ones to include.[35]

'Capsules' may offer a path for better aggregating natural language features into discernable packages. KBpedia's unique way of organizing and classifying related feature types based on the universal categories may also offer a better way to create meaningful 'capsules' for NLP. The 'capsule' approach, or other similar ways to package features into meaningful sets, may provide the missing technique for making deep learning more understandable in the context of natural language.

Measuring Classifier Performance

We presented statistical measures for binary classification and NLP tasks in *Chapter 14*. We touched upon but did not elaborate two additional measures of ROC and AUC. ROC, the receiver operating characteristic (also called the relative operating characteristic), is a curve that plots the true positive rate versus the false positive rate at various settings. AUC measures the area under this curve and reduces the standard error from the use of ROC alone. Researchers use these two measures to compare the performance of machine learning classifiers, though they are noisy methods with challenges in interpretation.[36] We need better performance measures.

In the 1930s the Italian statistician Bruno de Finetti wrote much on probability,[37] and likely was instrumental in resurrecting interest in Bayesian conditional probabilities. De Finetti developed a method of plotting three variables against one another called the ternary plot. It has found wide use in genetics, for example, in plotting the frequency of diploid genes (AA – Aa – aa) against one another using the display within an equilateral triangle, which can, for example, capture the distribution of the Hardy-Weinberg frequency of a gene, a standard measure.

About 15 years ago, the Spanish statistician Valverde-Albacete and his team adopted the de Finetti ternary plot to provide a more accurate means to compare machine learning classifiers. The plot uses the three corresponding values of change in entropy, versus what they termed the variation of information, and the mutual information surfaced by the classifier.[38] The group calls this display the 'entropy triangle.' One can see a striking resemblance of these de Finetti entropy triangles to the semiotic triangles of Peirce (see *Figure 2-1*). Further, the relation to Shannon entropy and the potential correspondences to object-representamen-interpretant at the apexes also draws attention. Though tentative, intuition about these correspondences suggests two possible lines of inquiry. First, we may apply de Finetti ternary plots to a more quantitative treatment of the Peircean sign representation. Second, the existing entropy calculations and insights might have either a Peircean interpretation or applicability to signs about Shannon information theoretics. For now, we should view these correspondences as wholly speculative, but thought-provoking nonetheless. Whether these intuitions bear fruit, the apparent superiority of the entropy triangle as a measure of classifier performance remains.

Thermodynamics of Representation

The close relation of information to energy as discussed in *Chapter 2* — and the findings of Landauer showing the energetic and physical aspects of information — provides possible guidance for how we should think about and model knowledge representations going forward. Susanne Still has taken this viewpoint to heart, and routinely uses the thermodynamic and informational aspects of information engines in her work.[39] This area, too, applies to measuring classifier performance, as well as other relevant topics.

For example, Still has shown information engines to require predictive inference to function well, which requires memory and favors a minimum of redundant information. In non-equilibrium conditions (namely, life), the most favored information engine is that which is most efficient in predictive power for a given level of memory. Of course, no information engine may extract more work than is contained in its useful informational inputs, and the best engines use more available information and dissipate less. (Dissipation under non-equilibrium conditions is average work minus the change in nonequilibrium free energy.) Still has also related her work to learning theory,[40] data representations, [41] and information bottlenecks.[42]

The idea of information bottlenecks to test for better data representations or better predictive inferences is but one method where we may exploit the convergence of information theory and knowledge. It is clear that we can apply these methods of entropy measurement to help screen data representations and models and even to test model parameters. These kinds of tests are hardly standard in ontology building and maintenance, though such efforts using the proxies of information engines provide a useful means for doing so.

We can apply these same perspectives and tests to evaluate the use of Peirce's universal categories as an organizational framework for knowledge graphs. We also should consider monitoring reference concepts by use to discover over-specified or redundant information in our systems. As was pioneered in biomedical research with 'knock-out' mice, we can remove selected pieces or portions of our knowledge graphs to measure their after and before information theoretic contributions.

As we pull together more evidence for the linkage between information theory and various entropy and free energy measures, we will undoubtedly discover more insights regarding composition and construction of our knowledge systems to make them more efficient. The beautiful thing about information-theoretic metrics is that we can negate empty arguments about philosophy or ideology and focus on what works with the most efficiency, a clear reflection of Peirce's admonitions for pragmatism. Routinely testing for information bottlenecks should also aid our ability to continue to refine better performing predictive inferences. Still states,

> "Predictive inference can be interpreted as a strategy for effective and efficient communication: past experiences are compressed into a representation that is maximally informative about future experiences. The information bottleneck (IB) framework can thus be applied, either in a direct way, or in its recursive form (RIB). Both methods find, asymptotically, the causal state partition, i.e., minimal sufficient statistics. RIB additionally recovers, asymptotically, the ε-machine, which is a maximally predictive and minimally complex deterministic HMM [hidden Markov model] , believed to be the best predictive description of a stochastic process that can be extracted from the data alone." (p. 985)

It appears pretty evident that we should adhere more to energetic factors (dissipation, entropy) in evaluating alternatives. These methods may also help us better quantify the benefits of organizing our knowledge structures using Peirce's universal categories and typologies as compared to traditional dichotomous representations.

POTENTIAL METHODS AND APPLICATIONS

New applications and uses for knowledge graphs remain untapped. We have listed some of these areas as potential applications for years, such as self-service business intelligence or semantic learning. We conclude this section and chapter by discussing the relation of Peirce's ideas and guidance to nature and questions of the natural world.

Self-Service Business Intelligence

I have been hearing about self-service business intelligence for more than two decades, yet it remains as elusive as ever. The definitions have changed over time and now include concepts like ‘big data,’ but the basic idea is to enable users, who lack IT or coding skills, to access enterprise data for their queries and reports.[43] The genesis of the idea arises from the promise of placing data analytics directly in the hands of the users who need it, matched with frustrations for how long specific requests to IT for queries or reports take to fulfill.[44] Part of the problem in achieving this ideal is parties tackled early attempts at self-service BI as some new application, only ‘dumbed down’ with slick user interfaces (UIs) to overcome the lack of computing skills by its users. In retrospect, it is not hard to see how attempts to fulfill this need settled upon supplying still another application as a separate product. Enterprise-level applications were the rage over those same decades. Naturally, to address the need of business analysts, the trick was to modify the business intelligence tools, such as they were, used by IT and then re-package them for easier use. The joke through at least the 1990s was that an ‘executive information system’ was the one with the big buttons with the big labels.

Those older visions fail for at least two reasons. The first reason is to consider business intelligence as some form of separate application. Early attempts at business intelligence or data warehousing failed and disappointed at high rates. We discussed at length in *Chapters 3* and *4* the challenges in data interoperability and impediments to information access and sharing. The general challenges of business intelligence and knowledge management remain unsolved. The second reason for failure is to consider the hurdle for non-technical users as mainly one of user interfaces. Sure, UI considerations are important. However, the real hurdles are fitting with existing work tasks and flows. The users of business intelligence create that intelligence. These knowledge workers must be involved in feeding and adding to the enterprise knowledge stores, as well as tapping them. Knowledge workers should steward their knowledge assets. This imperative needs to put users in the knowledge recording role, as well as the knowledge using one. Knowledge is not an afterthought, but part-and-parcel of the daily activities seamlessly integrated into current work tasks and flows.

Though KBpedia and its structure are well-suited to knowledge capture and use, the question of self-service goes beyond that. Self-service is not a matter of user interfaces and buttons, though at some point those items are worthy of attention, but a matter of mindset and making knowledge management integral to current work tasks. As we discussed in *Chapter 12*, this approach includes being attentive to workflows and piecing apart specific tasks such that they can integrate well with current daily activities (see further *Chapter 16*). As for knowledge creation, we must integrate new concepts and add and modify instances as we encounter them. These activities occur while researching online, writing or reading documents, or while interacting with co-workers and colleagues.

We need to deploy our specific KM apps where we engage in these activities — be they browsers, word processors, spreadsheets, calendar systems, or chat. BI systems would benefit from a similar distribution with standard work tasks. We want to encourage continuous access and constant availability. In these senses, we solve the UI challenge more by embedding knowledge functionality in existing applications than by dashboards or big buttons. While we have not been prescriptive in this book, I do think the guidelines we offer provide pragmatic approaches to adopt self-service ways in your organization.

Semantic Learning

Many aspects of what some anticipated as a semantic learning Web by 2020 have failed to materialize.[45] We have tried and used both latent and explicit ways to learn from text. We do not have multiple knowledge bases talking to each other or annotated or guided educational resources or commercial semantic browsers. We lack the connectedness portions of the vision. We have achieved talking to personal devices and leveraging massive knowledge bases like Wikipedia, mostly through supervised means, but the learning and interoperability aspects still appear weak. The lack of connection or connected learning sources is not one of technology or standards but provenance and authoritativeness. We have learned in our two decades of using the Web that it is a medium as prone to spamming and misuse as it is for access and convenience. We have found that the latent methods, applied to either text or images, do not perform as well as supervised methods. Still, though, even with supervised learning, we do not see much active learning or connectivity (defined as two separately maintained sources interacting automatically with one another).

We will not see marked improvements in latent semantic indexing -- and unsupervised methods in general -- until we have better parts-of-text segmentation and classification. We need a true foundational set of semantic primitives. I believe Peirce offers such (see next chapter). We have not yet tested this premise. Further, with its graph structures and inherent connectedness, we also have some exciting graph-learning methods that we can apply to KBpedia and its knowledge bases. The perhaps best-known method for conducting unsupervised learning on a sub-graph is the k-nearest neighbor method, with the latent Dirichlet allocation and conditional random fields (CRF) methods growing in popularity. We also have emerging sub-graph alternatives. With KBpedia's rich feature set, we have many additional options for discovering better-performing semantic learning. Whether the approach is Peircean or not, we likely need to see a more grounded set of semantic primitives emerge before we see production-grade performance with latent indexing or vectors. Without these primitives, there remains too much of a 'black box' aspect for these methods, similar to what we see with the opaque explanations for deep learning.

We do, however, have adequate means for production-grade methods for meaningful semantic connections using supervised learning with human editorial vetting. We need to take care of what resources we select for our learning purposes. We need automated ways to screen through the myriad of candidates. Then, we need to review those manually that remain ambiguous after tests, feeding our final selections back into the system to improve the performance of the learner when next used.

KBpedia thus is a potential contributor to semantic learning in two ways. The first way is to move toward a more logical, defensible set of semantic primitives for characterizing and indexing text, perhaps including unsupervised methods. The second way is by mindset and example, where builds and testing are constant against an already coherent structure. A key insight is in how to construct and maintain our knowledge structures. Users of the open Web, as is, do not trust it as a coherent knowledge source. Still, we will use quality sources, determined by editorial oversight or supplied by trusted brands. We need to discriminate and then depend on vetted resources, like from industry standards groups or proven resources like the Wikimedia properties. A key lesson is that we cannot fully automate the entire process of discovery, harvesting, vetting and connecting; humans must be in the loop, only accepting what meets editorial standards.

Nature As An Information Processor

It is clear that information is central to the idea of life (through DNA) and language and communications (through symbols). We also saw in the discussion in *Chapter 2* that Landauer had shown the physical nature of information and from Jaynes onward that many had pointed to the energetic nature of information. These indicators suggest that nature acts as an information processor.

The least controversial interpretation of information processing in nature occurs through genetic and cultural information. This overlap has led Sweller and Sweller to posit five common principals of natural information processing systems, which I have taken the liberty to edit slightly:[46]

Principles	*Cognitive Case*	*Evolutionary Case*	*Function*
Store information	Long-term memory	Genome	Store information for indefinite periods
Borrow and reorganize	Transfer information to long-term memory	Transfer information to the genome	Permit the rapid building of an information store
Random genesis	Create novel ideas	Create novel genetic codes	Create novel information
Narrow limits of change	Working memory	Epigenetic system related to environmental information	Input environmental information to the store
Organize and link	Long-term working memory	Epigenetic system related to genetic information	Use information from the information store

Table 15-1: Natural Information Processing System Principles

Wiesner — after reviewing developments in dynamical systems theory, information theory, physics, and computation theory — goes much further.[47] She claims that formal language theory, such as the examples of transformations provided by Noam Chomsky,[48] provides the key to understanding information processing in natural systems. Her synthesis leads to methods based on how quantum processes store and manipulate information, what Wiesner calls 'intrinsic quantum computation.'

In a broader sense, the mathematician Burgin and his co-authors over the years have been looking at commonalities and classification of various kinds of computational algorithms (for example, see[49]). Burgin claims the basic structure of the world is triadic (physical, structural, mental), which corresponds to Plato's triad (material, ideas/forms, mental) or may be related to Peirce's semiotic sign triad of object, sign, and interpretant. This existential triad leads Burgin and Dodig-Crnkovic to propose the three following types of computations:[49]

1. Physical or embodied (object) computations;
2. Abstract or structural (sign) computations; or
3. Mental or cognitive (interpretant) computations.

The authors note that the abstract or mental forms are themselves based on physical or embodied computations. In any case, the authors stress that we need a much better understanding of computation as an activity of information processing.

Quax, I believe, in his 2014 Ph.D. thesis[50] and associated papers, may have done just that. Returning to the roots of computation in Shannon information theory, as discussed in *Chapter 2*, Quax notes that the topological analysis of network interactions, while often posited as an explanatory basis, has proven insufficient to identify which nodes "drive the state" of networks.[51] Their idea, which supplements the topological relationships, is grounded in Shannon entropy and mutual information. Information theory is often applied to statistical inference when an external observer describes the state of a system. As applied to dynamical systems, such as knowledge systems, each component of the system (*e.g.*, a chunk of information) is an observer that stores the information and records state.

Quax and his co-authors derive two dynamic measures from these aspects of Shannon information. First, the authors calculate the influence of this information as it moves further from the source node, incurring losses on the way. They call this the 'information dissipation length.' (They measure IDL to the 50% dissipation level since the decay rate is asymptotic with a long tail.) IDL is a measure of the size of the subsystem that is affected by a particular element. IDL is somewhat akin to 'influence' in traditional graph measures that lack dynamic considerations (that is, are only topological). Second, the authors also calculate how the usefulness of the information dissipates over time. IDT is a proxy for how long the network remembers the particular state of a node, another measure of its influence.

This combination of structural (topographic) and dynamic (IDL and IDT) may not be exactly the right mix, but it does show how basing the analysis on information theoretics offers up new ways for understanding the nature of graphs and their interactions over time. For example, one finding is that it is intermediate players, not the central hubs or most popular nodes, that may have the most influence on dynamical processes within complex networks.[52] We may apply IDL and IDT to any complex, dynamic network. Mutual information (I) is that which nodes share. Here is a two-node example:

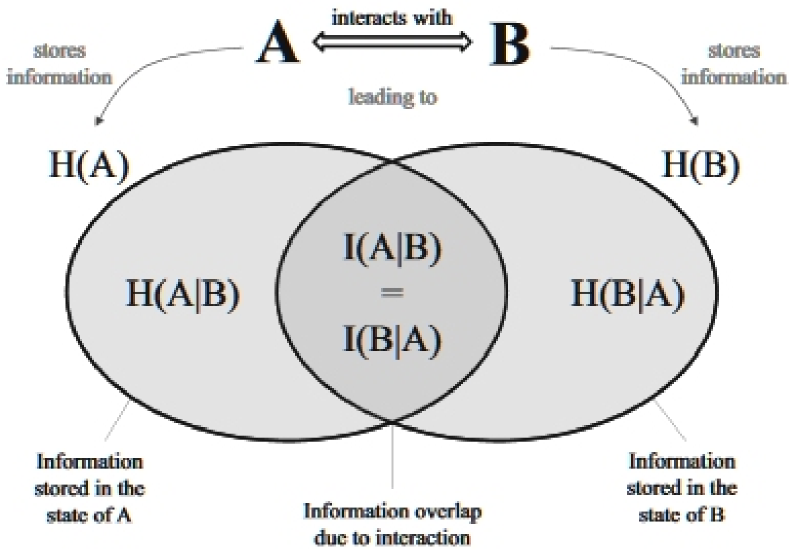

Figure 15-1: Shannon and Mutual Information (reprinted by permission of *The European Physical Journal – Special Topics*)[51]

'Deep information networks' use somewhat similar information-theoretic approaches to reduce the dimensionality of knowledge graphs,[53] though with potentially better understandability of the intermediate layers than deep learning. As we apply such techniques to more systems, we should gain further insights to improve our predictive power, perhaps getting to such seemingly intractable questions such as emergence, state transitions, or self-organization.

We see the potential relatedness or interactions between Peirce's semiosis, universal categories, and information theory. If we find that Peirce's universal categories indeed capture some fundamental truths about nature, for which some combination of the categories and information theory provides insight, then we can begin to apply lessons from natural science to the questions of language, knowledge, and representation. Each subsequent insight will feedback upon those that came before to improve our ability to model and predict our natural world.

Gaia Hypothesis Test

The chemist James Lovelock first posed the Gaia hypothesis* in the 1970s, soon

* Gaia was the Greek goddess who personified Earth.

getting collaborative support from the microbiologist Lynn Margulis.* They hypothesized that life is an integral part of the Earth's development. Organisms have co-evolved with changes in Earth geology and chemistry and climate; high oxygen levels, which are highly reductive, grew in the atmosphere due to the presence of life; life adapted to salinity changes due to salt run-off from terrestrial sources; and a complete weave of interacting forces and effects intertwined. The hypothesis has led some to consider the Earth a form of 'living thing.' Though derided when first postulated, advocates have refined the hypothesis to reflect emerging science better and scientists now largely embrace the idea of an evolving and interacting biosphere.

We also see another trend. The initial understanding of entropy as something that led to disorder caused thoughtful physicists, such as Erwin Schrödinger, discussed in *Chapter 2*, to posit explanations in the 1940s for how life did not violate the 2nd law of thermodynamics. That subject, too, has evolved much, whereas now a significant portion of scientists see entropy as operating in either equilibrium or non-equilibrium circumstances. The Earth, with massive influxes of solar radiation and the evolution of life that has created its 'Gaia-like' effects, is the quintessential non-equilibrium case.

Under non-equilibrium conditions with massive external influxes of energy, the equilibration principle, what one might also think of as selective pressure, is to dissipate this free energy as rapidly as possible. That idea, in turn, promoted on both statistical mechanics and biological terms by some, is known as the maximum entropy production (MEP) principle.[54] The principle favors structures that utilize and then dissipate free energy fastest and most efficiently. Ludwig Boltzmann, the explicator of entropy and statistical mechanics, is now praised by some for quantifying what is *not* (that is, entropy), akin to the contribution of the Arabian mathematicians who invented the number zero.[55]

Researchers have applied MEP to the Earth at planetary scale[56] and related it to more prosaic observations like water flows in soils.[57] Kleidon, in a comprehensive treatment of this topic with wonderful illustrations of various global fluxes, stated,"This seeming contradiction [of standard interpretations of entropy] is resolved by considering planet Earth as a coupled, hierarchical and evolving non-equilibrium thermodynamic system that has been substantially altered by the input of free energy generated by photosynthetic life." [58]

Herrmann-Pillath has woven these threads of the Gaia hypothesis, MEP, Charles Peirce's semiotics, and other factors into a complete speculation.[59] He includes the 'fourth law of thermodynamics' from Stuart Kauffman,[60] another theorist on the origin of life, who posed the role of work and the "tendency for self-construction biospheres to construct their own workspace." (p. 244) This view bridges from Peirce's statements about semiosis and its applicability to crystals and bees. We call the application to living organisms biosemiotics, and for inanimate or broader applications, such as what Herrmann-Pillath proposes, 'physiosemiosis.' This term arises from the proposition that "the biosphere is a system of generating, processing and storing information, thus directly treating information as a physical phenomenon," and fol-

* I discuss Margulis in a different context in *Chapter 3*.

lows the triadic semiotic model. A few researchers have speculated that Peirce's ideas of semiosis may even extend as far as the formation of matter after the Big Bang,[61] though it would be 15 years after Peirce's death before Hubble discovered the redshift. Still, Peirce intended his views on semiosis to infuse nature.

Peirce's advocacy that first, second, and third are the necessary and sufficient building blocks for all of reality may provide some missing insight into these basic questions of evolution and cosmology. His placement of randomness and chance into Firstness appears to conform with what we continue to learn about what is possible and where it arises. Peirce's prescience about signs, the universal categories, and the roles of chance and continuity quite possibly were truly cosmic. If indeed Peirce did grok the nature of nature at its most fundamental levels, then how we can apply his insights to our understanding of existence and reality is but at the beginning stages.

Chapter Notes

1. Xu, Y., Malt, B. C., and Srinivasan, M., "Evolution of Word Meanings Through Metaphorical Mapping: Systematicity Over the Past Millennium," *Cognitive Psychology*, vol. 96, Aug. 2017, pp. 41–53.
2. Ramiro, C., Malt, B. C., Srinivasan, M., and Xu, Y., "Mental Algorithms in the Historical Emergence of Word Meanings," *Cognitive Science*, 2017, pp. 986–991.
3. Oele, D., and Van Noord, G., "Distributional Lesk: Effective Knowledge-Based Word Sense Disambiguation," *IWCS 2017—12th International Conference on Computational Semantics—Short papers*, 2017.
4. Mikolov, T., Chen, K., Corrado, G., and Dean, J., "Efficient Estimation of Word Representations in Vector Space," *arXiv:1301.3781 [cs]*, Jan. 2013.
5. Yarowsky, D., "Unsupervised Word Sense Disambiguation Rivaling Supervised Methods," *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 1995, pp. 189–196.
6. Resnik, P. S., "Selection and Information: A Class-Based Approach to Lexical Relationships," Ph.D., University of Pennsylvania, 1993.
7. Sun, X., "Structure Regularization for Structured Prediction," *Advances in Neural Information Processing Systems*, 2014, pp. 2402–2410.
8. Banko, M., and Brill, E., "Scaling to Very Large Corpora for Natural Language Disambiguation," *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 2001, pp. 26–33.
9. Ponzetto, S. P., and Navigli, R., "Knowledge-Rich Word Sense Disambiguation Rivaling Supervised Systems," *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2010, pp. 1522–1531.
10. Agirre, E., Barrena, A., and Soroa, A., "Studying the Wikipedia Hyperlink Graph for Relatedness and Disambiguation," *arXiv:1503.01655 [cs]*, Mar. 2015.
11. Hearst, M. A., "Automatic Acquisition of Hyponyms from Large Text Corpora," *Proceedings of the 14th Conference on Computational Linguistics-Volume 2*, Association for Computational Linguistics, 1992, pp. 539–545.
12. Mintz, M., Bills, S., Snow, R., and Jurafsky, D., "Distant Supervision for Relation Extraction without Labeled Data," *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, Suntec, Singapore, 2-7: 2009, pp. 1003–1011.
13. Nguyen, D. B., Theobald, M., and Weikum, G., "J-REED: Joint Relation Extraction and Entity Disambiguation," Singapore, Singapore: ACM Press, 2017, pp. 2227–2230.
14. Bach, N., and Badaskar, S., "A Review of Relation Extraction," *Literature Review for Language and Statistics II*,

vol. 2, 2007, pp. 1–15.

15. Paulheim, H., “Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods,” *Semantic Web*, vol. 8, 2017, pp. 489–508.
16. Significant portions of this section are drawn from the KBpedia Web site for its reciprocal mapping use case; see http://kbpedia.com/use-cases/extending-kbpedia-with-kbpedia-categories/.
17. Perozzi, B., Al-Rfou, R., and Skiena, S., “DeepWalk: Online Learning of Social Representations,” *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 701–710.
18. See https://patents.google.com/patent/US8484245B2/en, “Large Scale Unsupervised Hierarchical Document Categorization Using Ontological Guidance,” Viet Ha-Thuc and Jean-Michel Renders, 2013, US Patent No. 8484245B2.
19. Ha-Thuc, V., and Renders, J.-M., “Large-Scale Hierarchical Text Classification Without Labelled Data,” ACM Press, 2011, pp. 685–696.
20. Traxl, D., Boers, N., and Kurths, J., “Deep Graphs - A General Framework to Represent and Analyze Heterogeneous Complex Systems Across Scales,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 26, Jun. 2016, pp. 1–27.
21. Johnston, B. G., and Williams, M., “A Formal Framework for the Symbol Grounding Problem,” *Conference on Artificial General Intelligence*, Atlantis Press, 2009.
22. Neches, R., Fikes, R. E., Finin, T., Gruber, T., Patil, R., Senator, T., and Swartout, W. R., “Enabling Technology for Knowledge Sharing,” *AI Magazine*, vol. 12, 1991, p. 36.
23. Song, Y., and Roth, D., “Machine Learning with World Knowledge: The Position and Survey,” *arXiv:1705.02908 [cs, stat]*, May 2017.
24. Cobreros, P., Égré, P., Ripley, D., and Rooij, R. van, “Foreword: Three-Valued Logics and Their Applications,” *Journal of Applied Non-Classical Logics*, vol. 24, 2014, pp. 1–11.
25. Lane, R., “Triadic Logic,” *Digital Encyclopedia of Charles S. Peirce*, 2001.
26. Elsenbroich, C., Kutz, O., and Sattler, U., “A Case for Abductive Reasoning Over Ontologies,” CEUR, 2006, pp. 1–12.
27. Du, J., Wang, K., and Shen, Y.-D., “A Tractable Approach to ABox Abduction over Description Logic Ontologies,” *AAAI*, 2014, pp. 1034–1040.
28. Prawat, R. S., “Dewey, Peirce, and the Learning Paradox,” *American Educational Research Journal*, vol. 36, Mar. 1999, pp. 47–76.
29. Flach, P., Kakas, A., and Ray, O., “Abduction, Induction, and the Logic of Scientific Knowledge Development,” *Workshop on Abduction and Induction in AI and Scientific Modelling*, Citeseer, 2006, p. 21.
30. Flach, P. A., “On the Logic of Hypothesis Generation,” *Abduction and Induction*, Springer, 2000, pp. 89–106.
31. Kapitan, T., “Peirce and the Autonomy of Abductive Reasoning,” *Erkenntnis*, vol. 37, 1992, pp. 1–26.
32. Kapitan, T., “In What Way Is Abductive Inference Creative?,” *Transactions of the Charles S. Peirce Society*, vol. 26, 1990, pp. 499–512.
33. Tschaepe, M., “Guessing and Abduction,” *Transactions of the Charles S. Peirce Society: A Quarterly Journal in American Philosophy*, vol. 50, 2014, pp. 115–138.
34. Hinton, G. E., Krizhevsky, A., and Wang, S. D., “Transforming Auto-Encoders,” *International Conference on Artificial Neural Networks*, Springer, 2011, pp. 44–51.
35. Sabour, S., Frosst, N., and Hinton, G. E., “Dynamic Routing Between Capsules,” *Advances in Neural Information Processing Systems*, 2017, pp. 3859–3869.
36. Valverde-Albacete, F. J., and Peláez-Moreno, C., “100% Classification Accuracy Considered Harmful: The Normalized Information Transfer Factor Explains the Accuracy Paradox,” *PLoS ONE*, vol. 9, Jan. 2014, pp. 1–10.
37. de Finetti, B., “Foresight: Its Logical Laws, Its Subjective Sources,” *Annales de l’Institut Henri Poincare*, vol. 7, 1937, pp. 94–158.

38. Valverde-Albacete, F. J., and Peláez-Moreno, C., "Two Information-Theoretic Tools to Assess the Performance of Multi-Class Classifiers," *Pattern Recognition Letters*, vol. 31, 2010, pp. 1665–1671.

39. Still, S., "Thermodynamic Cost and Benefit of Data Representations," *arXiv:1705.00612 [cond-mat]*, Apr. 2017, pp. 1–8.

40. Still, S., "Information Theoretic Approach to Interactive Learning," *EPL (Europhysics Letters)*, vol. 85, Jan. 2009, pp. 1–6.

41. Still, S., Sivak, D. A., Bell, A. J., and Crooks, G. E., "Thermodynamics of Prediction," *Physical Review Letters*, vol. 109, 2012, pp. 1–5.

42. Still, S., "Information Bottleneck Approach to Predictive Inference," *Entropy*, vol. 16, Feb. 2014, pp. 968–989.

43. TechTarget, "What Is Self-Service Business Intelligence (BI)?," *SearchBusinessAnalytics* Available: http://searchbusinessanalytics.techtarget.com/definition/self-service-business-intelligence-BI.

44. Stangarone, J., "Self-Service Business Intelligence 101," *mrc's Cup of Joe Blog*, Jul. 2015.

45. Stutt, A., and Motta, E., "Semantic Learning Webs," *Journal of Interactive Media in Education*, vol. 2004, 2004, pp. 1–32.

46. Sweller, J., and Sweller, S., "Natural Information Processing Systems," *Evolutionary Psychology*, vol. 4, Jan. 2006, p. 147470490600400130.

47. Wiesner, K., "Nature Computes: Information Processing in Quantum Dynamical Systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 20, Sep. 2010, p. 037114.

48. Chomsky, N., "Three Models for the Description of Language," *IRE Transactions on Information Theory*, vol. 2, 1956, pp. 113–124.

49. Burgin, M., and Dodig-Crnkovic, G., "Typologies of Computation and Computational Models," *arXiv preprint arXiv:1312.2447*, 2013, pp. 1–24.

50. Quax, R., "Information Processing in Complex Networks," Ph.D., University of Amsterdam, 2013.

51. Quax, R., Apolloni, A., and Sloot, P. M. A., "Towards Understanding the Behavior of Physical Systems Using Information Theory," *The European Physical Journal Special Topics*, vol. 222, Sep. 2013, pp. 1389–1401.

52. Quax, R., Apolloni, A., and Sloot, P. M. A., "The Diminishing Role of Hubs in Dynamical Processes on Complex Networks," *Journal of The Royal Society Interface*, vol. 10, Sep. 2013, pp. 1–10.

53. Franzese, G., and Visintin, M., "Deep Information Networks," *arXiv:1803.02251 [cs]*, Mar. 2018, pp. 1–10.

54. Kleidon, A., and Lorenz, R., eds., *Non-Equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond*, Berlin ; New York: Springer, 2005.

55. Ulanowicz, R. E., Goerner, S. J., Lietaer, B., and Gomez, R., "Quantifying Sustainability: Resilience, Efficiency and the Return of Information Theory," *Ecological Complexity*, vol. 6, Mar. 2009, pp. 27–36.

56. Kleidon, A., "Non-Equilibrium Thermodynamics, Maximum Entropy Production and Earth-System Evolution," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 368, Jan. 2010, pp. 181–196.

57. Zehe, E., Blume, T., and Bloschl, G., "The Principle of 'Maximum Energy Dissipation': A Novel Thermodynamic Perspective on Rapid Water Flow in Connected Soil Structures," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 365, May 2010, pp. 1377–1386.

58. Kleidon, A., "Life, Hierarchy, and the Thermodynamic Machinery of Planet Earth," *Physics of Life Reviews*, vol. 7, Dec. 2010, pp. 424–460.

59. Herrmann-Pillath, C., *Revisiting the Gaia Hypothesis: Maximum Entropy, Kauffman's 'Fourth Law'and Physiosemeiosis*, Frankfurt School of Finance & Management, 2011.

60. Kauffman, S. A., *Investigations*, Oxford University Press, 2000.

61. See Note 19 in *Chapter 10*.