Appendix A: Perspectives on Peirce

Charles Sanders Peirce (1839–1914), pronounced 'purse,' was an American logician, scientist, mathematician, and philosopher of the first rank. His profound insights and writings spanned a half-century, and cover topics ranging from the nature of knowledge and epistemology to metaphysics and cosmology. Well known in the Americas and Europe early in his career, his foibles, and challenges to social orthodoxies, led to a precipitous decline in his fortunes, such that he died nearly penniless and unable to publish. Still, Peirce had a deep influence on many leading thinkers of his time, and as transcription and publishing of his voluminous writings move to completion, an influence that will continue for generations.

My first attraction to Peirce began with my professional interests in the semantic Web.[1] My earliest exposure to the semantic Web kept drawing my attention to questions of symbolic knowledge representation (KR). Like the genetic language of DNA in biology, my thought has been that there must be better (more 'truthful') ways of representing knowledge and information in digital form. My sense is that syntax or specific language is not the key, but that the basic building blocks of grammar and primitives hold that key. We further need a set of primitives well suited to natural language understanding, since humanity embodies so much of its cultural information in text. Structured data, such as from databases, is not an appropriate starting point; we critically need means to represent natural language. In Peirce, I have found the guide for those interests.

I have maintained throughout this book that Peirce is the greatest thinker ever in the realm of knowledge representation. Yet, KR, as a term of art, was not a phrase used in Peirce's time. Granted, Peirce wrote much on relations and representation (via his semiotic theory of signs) and provided many insights into the nature of

[1] Some material in this appendix was drawn from the author's prior articles at the *AI3:::Adaptive Information* blog: "The Importance of Being Peirce" (Sep 2016); "Being Informed by Peirce" (Feb 2017); "How I Study C.S. Peirce" (Aug 2017); "Why I Study C.S. Peirce" (Aug 2017); "A Foundational Mindset: Firstness, Secondness, Thirdness" (Mar 2016); "How I Interpret C.S. Peirce" (Sep 2017).

M. K. Bergman, *A Knowledge Representation Practionary*,
https://doi.org/10.1007/978-3-319-98092-8

information and knowledge, but he never used the specific phrase of 'knowledge representation.' He never attempted to categorize knowledge such as what we have undertaken with the KBpedia Knowledge Ontology (KKO), though he did make multiple attempts to classify the 'sciences' (fields of study in today's parlance). While Peirce had more than a glimmer of an idea that reasoning machines might someday be a reality, there was no need within his time to attempt to provide the specific representational framework for doing so.

Because of his influence—and his nearly constant presence throughout this book—I wanted to share what I have learned about Peirce the person, the polymath, the philosopher, and as a polestar guiding new directions in KR. I hope to convey a bit of the perspective about why you, too, should study Peirce, and help add to the interpretation of his fecund mind. I conclude this appendix with suggested resources you may find helpful to study this most remarkable human thinker.

Peirce, the Person

Charles S. Peirce was born into privilege in 1839 and was brought up among the intellectual elite in Cambridge, Massachusetts. His father, Benjamin Peirce, was a professor at Harvard and one of the prominent mathematicians of the 1800s. Charles received a first-rate education, including much personal tutoring by his father, and was given preference and sinecures at a young age, mainly through his father's connections.

Trained as a chemist at Harvard's Lawrence Scientific School where he graduated *summa cum laude* in 1863, Peirce was able to secure a deferment with his father's assistance from serving in the Civil War. Peirce was a working scientist for most of his employed career at the Coast and Geodetic Survey, then perhaps the premier US Government research facility, on gravitational differences around the globe, based on meticulous measurements using pendulums, often of Peirce's innovative designs. His early writings in the mid-1860s in areas of logic and metaphysics received wide acclaim. He was frustrated in securing a teaching position at Harvard,[2] but eventually became a lecturer at Johns Hopkins University, which was innovating in American graduate education, from 1879 to 1884, when he was summarily dismissed under unclear clouds of scandal. He subsequently had sporadic engagements in various entrepreneurial activities and wrote and translated articles for hire, but never had a permanent position again. His last decades were spent writing at his Milford, Pennsylvania home, Arisbe, which itself was in various stages of construction and disrepair based on vacillating, but declining, financial fortunes. By his death in 1914, he and his second wife, Juliette, were essentially penniless, having been sustained in part due to loans and charity from friends and family, orchestrated by his brother, James, himself a Harvard mathematician, and his lifelong friend, William James. Peirce had no children.

[2] In fact, due to enmity at Harvard, Peirce was barred from lecturing on campus for 30 years, only relaxed when Peirce was in his 60s.

In a stellar biography, Brent often refers to Peirce as a dandy in his earlier years [1]. Playing on the pronunciation of his name, two of Peirce's favorite self-descriptions were that he had 'Peirce-istence' and 'Peirce-everence.' He was certainly an iconoclast, and also flaunted society's conventions, living with Juliette before marriage and after being abandoned by his first wife, Zina. Peirce was a prodigious writer and very hard worker over 50 years, but was cavalier, if not unethical, in the abuse of his positions and public funds. He was reportedly a user of morphine and cocaine, ostensibly for neuralgia, but a factor that may have contributed to his sometimes perplexing inconstancies. Peirce often pursued his intellectual interests at the expense of his paid responsibilities. He created powerful enemies that ultimately kept him from securing a professorship at a leading university, which he and his family believed his birthright. He made poor decisions concerning money and finances, often disastrous ones, and died virtually penniless, with no fame and little notoriety. Still, Peirce befriended and influenced many of the leading thinkers of his time, including William James, Josiah Royce, John Dewey, and Oliver Wendell Holmes.

After Peirce's death, Harvard was scandalous in how it (mis-) handled his donated papers and restricted access for many years to his unpublished writings,[3] a continuation of the vendetta brought by Charles W. Eliot, the long-standing Harvard president. His supposed supporter and family friend, Simon Newcomb, routinely undercut Peirce. Thankfully, within two decades of his death, anthologies were published, and his reputation and stature began to grow. The understanding of his insights and accomplishments continues to grow as researchers study and release his voluminous unpublished writings. Peirce's reputation now is the highest it has ever been in the 100 years since his death, growing, and surely greatly exceeds whatever fame he saw during life.

Peirce was often the first to acknowledge how he changed his views, with one set of quotes from early 1908 showing how his thinking about the nature of signs had changed over the prior 2 or 3 years [3]. That example is but a small snapshot of the changes Peirce made to his sign theories over time, or of his acknowledgments that his views on one matter or another had changed.

Of course, it is not surprising that an active writing career, often encompassing many drafts, conducted over a half of a century, would see changes and evolution in thinking.[4] Most Peircean scholars acknowledge changes in Peirce's views over time, particularly from his early writings in the 1860s to those after the turn of the century and up until his death in 1914. Where Peirce did undergo major changes or refinements in understanding, Peirce himself was often the first to explain those changes. Many scholars have looked to specific papers or events to understand this evolution in thinking. Max Fisch divided Peirce's philosophy development into three periods: (1) the Cambridge period (1851–1870); (2) the cosmopolitan period (1870–1887);

[3] See further Nathan Houser, "The Fortunes and Misfortunes of the Peirce Papers" [2].

[4] Peirce's lifetime writings have been estimated at 100,000 pages, and Case has estimated that as many as three-quarters of his writings still wait transcription [4]. I doubt this estimate, but in any case discovery of new entire manuscripts is unlikely, since untranscribed pages seem to constitute mostly drafts of prior manuscripts.

and (3) the Arisbe period (1887–1914) [5]. Murphey split Peirce's development into four phases: (1) the Kantian phase (1857–1866); (2) three syllogistic figures (1867–1870); (3) the logic of relations (1879–1884); and (4) quantification and set theory (1884–1914) [6]. Brent has a different split more akin to Peirce's external and economic fortunes [1]. Parker tends to split his analysis of Peirce into early and mature phases [7]. It is a common theme of major scholars of Peirce to note these various changes and evolutions. Some of this analysis asserts breakpoints and real transitions in Peirce's thinking. Others tend to see a more gradual evolution or maturation of thinking. Some of the arguments bolster whatever particular thesis the author is putting forward. Such is the nature of scholarship.

For me, I take a pragmatic view of these changes. First, some of Peirce's earliest writings, particularly his 1867 "On a New List of Categories' [8], but also mid-career ones, are amazingly insightful and thought provoking. Tremendous value resides in these earlier writings, often infused with genius. Peirce, after all, was in the prime of his powers. Sure, I can see where some points have evolved, or prior assertions or terminology have changed, but Peirce is also good at flagging those areas he sees as having been important and earlier in error. I, therefore, tend to rely most on his later writings, when a hard life lived, maturity and experience added wisdom and perspective to his thoughts. I tend to see his later changes more as nuanced or mature, rather than radical breaks with prior writings. I see tremendous continuity and consistency of worldview in Peirce over time.[5]

Peirce considered himself foremost as a scientist, who probed and questioned premises with logic and purpose. Peirce's critical attention and refinement of the scientific method place him in the top tier of philosophers of science. Peirce believed that all questions lend themselves to scrutiny and logical analysis. Among the myriad of possibilities available to us for inquiry as scientists, Peirce's methods help point to those options most likely to yield fruit within limited time and resources, the essence of his philosophy of pragmatism. The universal categories provide us with a constant and consistent framework for representing, analyzing, and organizing knowledge.

Though many intellectual giants of history were recognized as such in their own times—Newton, Einstein, Darwin, and Aristotle come to mind—all of us like the story of the genius unjustly ignored in his lifetime. In science, famous examples include Copernicus, Galileo, Wegener, and Mendel. Charles Sanders Peirce belongs in this pantheon as well, a possible outcome I think he realized himself. The failure of his grant application to the Carnegie Institution in 1903 to synthesize his life's work, supported no less by Andrew Carnegie and Teddy Roosevelt, was Peirce's last attempt at broad-scale recognition. Ill, in poverty, and shunned by the establishment of his time, Peirce worked feverishly in his last years to get down on paper as much

---

[5] I make this assertion despite major shifts in some of Peirce's positions. For instance, Max Fisch in his 1967 paper, "Peirce's Progress from Nominalism Toward Realism," in *The Monist* (vol. 51, pp. 159–178) charts Peirce's evolution from some nominalist positions in his early writings to a full-blown "three-category realist" (Fisch's phrase) by the turn of the century. Still, this evolution—and others—only augments Peirce's lifelong theses regarding sign-making, logic, and universal categories.

as he could, pretty much laboring alone and in obscurity. We are still plumbing these handwritten papers, gaining new insights and perspectives of what we think we know about Peirce's philosophy and perspectives.

Philosophers, logicians, scholars, and laypersons study Peirce as a passion, many for a living. Though Peirce was neglected by many during the heyday of analytical philosophy throughout the twentieth century, that is rapidly changing. Walker Percy and Umberto Eco were two noted writers who have studied Peirce closely and written on him. The reason for Peirce's ascendancy, I think, is precisely due to the Internet, with then ties to knowledge representation and artificial intelligence. Peircean views are directly relevant to those topics. His writings in logic, semiosis (signs), pragmatics, existential graphs, classification, and how to classify are among the most direct of this relevancy.

However, relevant does not mean agreed upon, and researchers understand Peirce through their own lenses, as the idea of Peirce's Thirdness affirms. Given Peirce's own constant questioning and revision of his theories, plus the fragmented nature of the written record he left behind, I think it is fair to assert that we will never come to understand Peirce's 'truth' fully. Peirce was a man of complexity, unlikely to be fully plumbed. On the other hand, I also think we are only still beginning to understand how Peirce's insights can inform our understanding of the world.

Peirce, the Philosopher

Peirce did not view himself as a philosopher but as a scientist and logician. These distinctions are mere shadings in Peirce's philosophy, one that places high stock on truth, logic, representation, and the scientific method. Much of Peirce's philosophy figures prominently in the main body of this book, specifically in the role of semiosis and sign-making (Chap. 2); his universal categories, "truth" and fallibility, and categorization (Chap. 6); logic of relations and logic types (Chaps. 7 and 8); the role of natural classes (Chaps. 5 and 10); and pragmatism (Chap. 14). Here, however, I want to highlight the more cross-cutting aspects of Peirce's philosophy, not so directly related to KR, but also essential to understand his worldview.

Peirce's Architectonic

Peirce's *architectonic*, a word applied to the worldview for certain influential philosophers such as Kant or Aristotle, is built around the structure of all human knowledge. The pivotal elements of Peirce's architectonic are his universal categories, as manifested in logic, and evaluated through the pragmatic maxim. Peirce organized his classifications of science into disciplines using this system, in which he also embedded such topics as ethics, esthetics (his spelling), philosophy, and metaphysics, in addition to the classical sciences and humanities. Peirce evolved his classification of the sciences considerably over time. Beverly Kent conducted a thorough analysis in 1987, much based on unpublished manuscripts at the time, that documents at least 20 different classifications over the period of 1866 to 1903 (the last, final one called the 'perennial'), with minor ones in between [9]. The three main branches of Peirce's perennial classification are mathematics, cenoscopy (philosophy), and idioscopy (the special sciences of traditional science and the humanities). Peirce believed that we should place philosophy within this systematic account of knowledge as science, and adopted the idea of the architectonic from the philosopher he idealized the most, Immanuel Kant. Peirce increasingly relied on this structural sense and the irreducible universal categories in most all of his later thinking.

Logic, as defined by Peirce, is only another name for semiotic (1897, CP 2.227). The clear thread through Peirce's writings is the respect and attention he gives to the primacy of logic, but also the role of community in deciding belief and terminology. Though, as a normative science (along with ethics and esthetics), logic is not the center root of his categorization of science, Peirce still bases all of his major arguments and insights on logic. Those insights include ones about the role and principles of logic itself.

> I do not, for my part, regard the usages of language as forming a satisfactory basis for logical doctrine. Logic, for me, is the study of the essential conditions to which signs must conform in order to function as such. How the constitution of the human mind may compel men to think is not the question; and the appeal to language appears to me to be made (and logicians generally do make it; in particular their doctrine of the copula appears to rest solely upon this), it would seem they ought to survey human languages generally and not confine themselves to the small and extremely peculiar group of Aryan speech (1904, NEM 4:243).

Via the classification of the sciences, Peirce attempts to organize and relate all aspects of knowledge and inquiry, and via the logic of semiosis Peirce provides a way to think about and represent that knowledge. Peirce subsumes these considerations under the irreducible foundation of the universal categories, though Peirce placed the study of these categories within phenomenology, another branch of philosophy. (Thus, phenomenology, normative science, and metaphysics provide the three branches of cenoscopy, or philosophy.) Peirce is also clear about these same groundings for his pragmatism. In a 1902 letter to William James, Peirce stated:

> [M]y three categories, … in their psychological aspect, appear as Feeling, Reaction, Thought. I have advanced my understanding of these categories much since Cambridge days; and can now put them in a much clearer light and more convincingly. The true nature of pragmatism cannot be understood without them (1902, CP 8.256).

Though we can see the universal categories subsuming logic, semiosis, and pragmatism, we can also see a tight nexus between all of the concepts. For example, Ika, in an overlooked doctoral thesis, provides lengthy analysis that places Peirce's universal categories at the foundation of his pragmatism [10]:

> … it can be said that Peirce's general philosophical project was most fundamentally concerned with some kind of *methodological* quest; a quest that seeks to establish the most fundamental categories that are both logically and metaphysically presupposed in any inquiry. The categories are logical presuppositions in the sense that they are principles or norms to be necessarily followed in the process of inquiry. They are also metaphysical presuppositions in the sense that Peirce rightly regarded them as reflections or representa-

> tions of reality. Peirce's unique brand of pragmatism, with its blend of logical rigour, practical orientation and realist metaphysical foundations was the end result of his methodological quest (p. 23).

which also ties into the idea and importance of logic:

> According to his classification of the sciences, metaphysics depends on logic for its fundamental principles, and logic depends on metaphysics for the data on which to operate. Although this relation of inter-dependence between metaphysics and logic is useful for determining certain aspects of his overall philosophical position, it is too rigid to account for another sense in which logic is dependent on metaphysics for Peirce, namely, that the whole end or intention of logic is contained within metaphysics [10] (p. 139).

but also recognizes that the categories subsume semiosis, providing the more general tenets:

> While Peirce appears to be preoccupied with his theory of signs, and sees sign and sign-action in every phenomenon, he did not seek to reduce reality to a semiotic system, where the real would be construed as only that which is sign-like. For Peirce, such a reductionist view of reality would result either in a dismissal of metaphysics or require that metaphysics be reducible to logic … both these views are inconsistent with Peirce's overall philosophical position, which recognises the distinction between the logical and the real as important [10] (p. 152).

We stride into a world with an uncertain future. We need to act and make decisions in the face of that uncertainty. We evaluate that world by the three logical methods of deduction, induction, and abduction. Peirce's architectonic provides a nexus of logic, signs, and universal categories to give us the tools we need to move forward, what Peirce calls pragmatism:

> Pragmatism … had been designed and constructed … architectonically. Just as a civil engineer, before erecting a bridge, a ship, or a house, will think of the different properties of all materials, and will use no iron, stone, or cement, that has not been subjected to tests; and will put them together in ways minutely considered, so, in constructing the doctrine of pragmatism the properties of all indecomposable concepts were examined and the ways in which they could be compounded. Then the purpose of the proposed doctrine having been analyzed, it was constructed out of the appropriate concepts so as to fulfill that purpose. In this way, the truth of it was proved (1905, CP 5.5).

We can thus understand Peirce's *architectonic* as the building blocks that go into constructing our structure of knowledge. How we go about thinking about these building blocks and then applying them to a given problem at hand, such as capturing a domain or inquiring where we have doubt, is what I refer to as a *mindset*. The universal categories are foundational to either of these two meanings.

Chance, Existents, and Continuity: Real

The three universal categories, as noted, are appropriately studied under the phenomenology section of the cenoscopic (philosophic) branch of the sciences. Though we earlier, in *Table 6.2*, listed many examples of Firstness, Secondness, and Thirdness, let's single out some phenomenological aspects of these categories that Peirce emphasized in his writings. These three aspects are absolute chance for Firstness; actual, existing individuals for Secondness; and continuity for Thirdness. In some ways, these concepts are firsts among equals given their prominence in Peirce's thinking. If a grounding exists for the three universal categories, these may be it.

Chance

The fount of Peirce's universal category of Firstness is absolute chance. Peirce brings two remarkable insights about chance in his writings. The first insight, now somewhat prosaic but new for its time, was the importance of probability to many problems. The results, for many problems, are not absolute, but probable across a distribution of possible outcomes. It is essential to sample randomly, or by chance, to test these probabilities. Peirce was an early explicator about random sampling and statistics. Indeterminant problems are common, and an understanding of chance and probabilities is the only tractable way to assess them.

The second remarkable insight is more fundamental, and perhaps even more critical. It is what Peirce called *tychism*. Peirce was an early supporter of Darwin's theory of evolution and understood the role of variation. Peirce was first exposed to the ideas of evolution at least since the time of the Metaphysical Club, under the strong influence of friend and fellow club member, Chauncey Wright. Peirce's probability studies also enabled him to see that our world was one of 'surprising facts.' A completely random world would signal no variety. Absolute chance must, therefore, be leading to variants that cause us to inspect and understand emerging properties. Chance is itself offering up variants, some which have the character of persistence because of their stronger probability to be reinforced. These forces of chance give our world the variety and diversity it possesses. Laws and habits lead to regularities that both tend to perpetuate themselves as generalities, but also flash surprising variation that causes us to take stock and categorize and generalize anew.

In Peirce's cosmogony, these primitives of chance (Firstness), law (Secondness), and habit (Thirdness) can explain everything from the emergence of time and space to the emergence of matter, life, and then cognition. Though it is true that Thirdness (continuity) is the more synthesizing concept, the role of chance alone to drive this entire reality suggests its essential character [11]. Tychism is thus a philosophical doctrine that absolute chance is real and operative in the world, and it is the source of irregularity and variety and the underlying force of evolution.

Chance alone could be the variant that led to the minute differences arising during the Big Bang, which is posited to have led to matter and its structure. Chance is what enabled self-perpetuating life to emerge from inanimate matter. Chance is how forms of life could symbolically capture these variations via cognition and language. While all of this may now seem inevitable—though unexpected in how manifested—Peirce would maintain that they are events arising from chance. Perhaps most events have a cause, but the fundamental ones result from chance. 'Surprising facts,' a favorite phrase of Peirce, mean the world is unpredictable and ultimately probabilistic. The limits of Cartesian logic, the 0's and 1's, are likely never achievable. Reality is shaded and nuanced.

When Peirce began putting forth these ideas, specifically in his *Popular Science* series in 1878 in "On the Nature" [12], these were radical ideas. At the time of these publications, science was still decades away from quantum mechanics and the Heisenberg uncertainty principle. Moreover, even though Einstein (in) famously said that "God doesn't play dice with the world,"[6] Einstein himself, and his unsettling of Newtonian physics, was still three decades away. These examples are but a few of where Peirce had insight and prescience well in advance of later supporting science.

The reason for such insights, Peirce would say, and I would agree, is not that he was somehow miraculously able to see the future. But, through the rigorous application of logic, Peirce was able to see the requisite primitives of existence. As he wrote,

> The endless variety in the world has not been created by law. It is not of the nature of uniformity to originate variation, nor of law to beget circumstance. When we gaze upon the multifariousness of nature we are looking straight into the face of a living spontaneity. A day's ramble in the country ought to bring that home to us (1887, CP 6.553).

Peirce posited five reasons to believe in the reality (objective existence) of absolute chance: [6] (1) mechanical forces cannot explain growth and complexity in nature; (2) the sheer variety of nature; (3) uniformity develops from some state of determinacy; (4) no empirical evidence supports determinism; and (5) we can draw verifiable consequences from the hypothesis of chance (from 1892, CP 6.58-62).

Existents

Existents are what is actual, what exists, and consists of events and entities. Everything that exists is an individual and has an identity. Existents reside entirely in Secondness. "… existence (not reality) and individuality are essentially the same thing" (1901, CP 3.613). Existents have the nature of 'haecceity,' the idea of 'thisness' from the Latin, that gives them their particular uniqueness and identity.

Existents are thus an instantiation, something actual with identity, in comparison to the possibilities or qualities of Firstness, and in contrast to the generalities or continuities of Thirdness.[7] Existents embody qualities as found in Firstness, and may be generalized or related to continuous collections as found in Thirdness. Existents have some limits that bound their thisness, or haecceity, in either space (entities) or time (events). They exist whether we think them to do so or not. In Peirce's semiosis, actual existents are sinsigns.[8] We may indicate

[6] See https://en.wikiquote.org/wiki/Albert_Einstein

[7] Peirce sometimes also refers to relations between two existent objects as also being *existent*.

[8] Jon Alan Schmidt takes exception to this wording, noting that it entails that everything exists is a sign, which perhaps Peirce never stated in exactly this way. The quote we do have from Peirce is "that all this universe is perfused with signs, if it is not composed exclusively of signs" (1906, EP 2:394).

existents (via an index) as an object. Existents are real, but reality is not limited to them. Secondness is the most straightforward of the universal categories.

Continuity

Synechism, which Peirce equated to continuity,[9] is the notion that space, time, and law are continuous and form an essential Thirdness of reality in contrast to existing things and possibilities. Peirce notes that continuity is one of "the most difficult, the most important, the most worth study of all philosophical ideas" (1893, MS 717; NEM 4:310). I tend to agree.

> Now if we are to accept the common sense idea of continuity (after correcting its vagueness and fixing it to mean something) we must either say that a continuous line contains no points or we must say that the principle of excluded middle does not hold of these points. The principle of excluded middle only applies to an individual (for it is not true that 'Any man is wise' nor that 'Any man is not wise'). But places, being mere possibles without actual existence, are not individuals. Hence a point or indivisible place really does not exist unless there actually be something there to mark it, which, if there is, interrupts the continuity… On the whole, therefore, I think we must say that continuity is the relation of the parts of an unbroken space or time… The precise definition is still in doubt; but Kant's definition, that a continuum is that of which every part has itself parts of the same kind, seems to be correct. This must not be confounded (as Kant himself confounded it) with infinite divisibility, but implies that a line, for example, contains no points until the continuity is broken by marking the points. In accordance with this it seems necessary to say that a continuum, where it is continuous and unbroken, contains no definite parts; that its parts are created in the act of defining them and the precise definition of them breaks the continuity… Breaking grains of sand more and more will only make the sand more broken. It will not weld the grains into unbroken continuity (1902, CP 6.168).

Peirce clearly excludes individuals from continuity; indeed, they are disruptions to it. The principle of excluded middle also does not apply, since we are also dealing with generalities. He illustrates these ideas in multiple passages with the idea of points on a continuous line, such as this next example:

> A true continuum is something whose possibilities of determination no multitude of individuals can exhaust. Thus, no collection of points placed upon a truly continuous line can fill the line so as to leave no room for others, although that collection had a point for every value towards which numbers, endlessly continued into the decimal places, could approximate; nor if it contained a point for every possible permutation of all such values. It would be in the general spirit of synechism to hold that time ought to be supposed truly continuous in that sense (1902, CP 6.170).

[9] Peirce states, "I have proposed to make *synechism* mean the tendency to regard everything as continuous" (1893, CP 7.565). He goes on to say, "I carry the doctrine so far as to maintain that continuity governs the whole domain of experience in every element of it. Accordingly, every proposition, except so far as it relates to an unattainable limit of experience (which I call the Absolute,) is to be taken with an indefinite qualification; for a proposition which has no relation whatever to experience is devoid of all meaning" (CP 7.566).

We cannot distinguish things without making the line discontinuous. If something is inexplicable, it cannot be continuous:

> … synechism amounts to the principle that inexplicabilities are not to be considered as possible explanations; that whatever is supposed to be ultimate is supposed to be inexplicable; that continuity is the absence of ultimate parts in that which is divisible; and that the form under which alone anything can be understood is the form of generality, which is the same thing as continuity (1902, CP 6.173).

We now begin to see the intimate connection between continuity and generality. "True generality is, in fact, nothing but a rudimentary form of true continuity. Continuity is nothing but perfect generality of a law of relationship" (1902, CP 6.172). We can also relate continuity to the concepts of regularity:

> That continuity is only a variation of regularity, or, if we please so to regard it, that regularity is only a special case of continuity, will appear below, when we come to analyze the conception of continuity. It is already quite plain that any continuum we can think of is perfectly regular in its way as far as its continuity extends. No doubt, a line may be say an arc of a circle up to a certain point and beyond that point it may be straight. Then it is in one sense continuous and without a break, while in another sense, it does not all follow one law. But in so far as it is continuous, it everywhere follows a law; that is, the same thing is true of every portion of it; while in the sense in which it is irregular its continuity is broken. In short, the idea of continuity is the idea of a homogeneity, or sameness, which is a regularity. On the other hand, just as a continuous line is one which affords room for any multitude of points, no matter how great, so all regularity affords scope for any multitude of variant particulars; so that the idea [of] continuity is an extension of the idea of regularity. Regularity implies generality; and generality is an intellectual relation essentially the same as significance, as is shown by the contention of the nominalists that all generals are names. Even if generals have a being independent of actual thought, their being consists in their being possible objects of thought whereby particulars can be thought. Now that which brings another thing before the mind is a representation; so that generality and regularity are essentially the same as significance. Thus, continuity, regularity, and significance are essentially the same idea with merely subsidiary differences. That this element is found in experience is shown by the fact that all experience involves time. Now the flow of time is conceived as continuous. No matter whether this continuity is a datum of sense, or a quasi-hypothesis imported by the mind into experience, or even an illusion; in any case it remains a direct experience. For experience is not what analysis discovers but the raw material upon which analysis works. This element then is an element of direct experience (1908, CP 7.535).

At one point, Peirce claims that "continuity represents Thirdness almost to perfection" (CP 1.337), and Haack notes that abductive reasoning is the preferred logic for positing continuities [13]. Peirce relates his concept of time to continuity (CP 6.132), and claims that his ideas about fallibility are grounded in it:

> The principle of continuity is the idea of fallibilism objectified. For fallibilism is the doctrine that our knowledge is never absolute but always swims, as it were, in a continuum of uncertainty and of indeterminacy. Now the doctrine of continuity is that all things so swim in continua (1897, CP 1.171).

Peirce notes that classifying and typing things are also grounded in continuity:

> … it will be found everywhere that the idea of continuity is a powerful aid to the formation of true and fruitful conceptions. By means of it, the greatest differences are broken down and resolved into differences of degree, and the incessant application of it is of the greatest value in broadening our conceptions (1878, CP 2.645).

We thus see that Peirce's conception of continuity is a metaphysical theory as well as a methodological principle. Peirce and others have noted that the presence of continuity is not a construct of the human mind, but is part of reality [14].

What Is Real

Peirce grew over his working life to believe that all of these universal categories were real, and not merely figments of the human mind. "If I *truly know* anything, that which I know must be *real*" (EP 2:181). Fisch dated this transition to about 1897 [15] when Peirce accepted the reality of the category of Firstness, *i.e.*, of possibility, in addition to his then acceptance of the reality of the categories of Thirdness and Secondness, becoming what Fisch called a 'three-category realist.'[10] In Peirce's words:

> 'Truth is the conformity of a representation to its object,' says Kant. One might make this statement more explicit; but for our present purpose it may pass. It is nearly correct, so far as it is intelligible. Only, what is that 'object' which serves to define truth? Why it is the *reality*: it is of such a nature as to be independent of representations of it, so that, taking any individual sign or any individual collection of signs (such, for example, as all the ideas that ever enter into a given man's head) there is some character which that thing possesses, whether that sign or any of the signs of that collection represents the thing as possessing that character or not. Very good: now only tell me what it means to say that an object possesses a character, and I shall be satisfied. But even now, in advance of our study of definition, [we can] sufficiently see that we can only reach a conception of the less known through the more known, and that consequently the only meaning which we can attach to the phrase that a thing 'has a character' is that something is *true* of it. So there we are, after threading the passages of this labyrinth, already thrown out at that very conception of truth at which we entered it. Indeed, when one comes to consider it, how futile it was to imagine that we were to clear up the idea of *truth* by the more occult idea of *reality*! (1902, CP 1.578)

Reality, for Peirce, is that which has character independent of what we might think about it in our minds. It rejects the Cartesian mind-body duality. The measures of a character of a thing arise from its disruptions in continuity. These disruptions arise from evolving design and absolute chance.

Leaning into Pragmatism

In a probabilistic world, which it is, we see lines of evidence everywhere for inferring various aspects of the world, now and into the future. The truth is, as Peirce often makes clear, only the here and now are knowable; what might come next (into the future) is a probability. The stronger, or more definitive, means of inference, deduction, and induction can never apply to the future. I am not sure Peirce understood that his formulation of abductive reasoning was the needed pathway here, but it is also true that abductive reasoning is the only path to new knowledge or novelty. We must make practical choices in our limited time. From Ika's dissertation: [10].

[10] h/t to Jon Alan Schmidt; also see EP 2:186-195.

> The point is that pragmatism as a logical maxim is set to serve the assertion that there are real things; for without that assertion, pragmatism would be a meaningless enterprise, no matter how hard we think of it as only a logical principle. In his classification of the sciences, Peirce describes logic as the science of the category of Secondness, and metaphysics as the science of Thirdness. His whole point is that the sciences are just as closely related to one another as are the three categories. That is, according to his theory of categories, Secondness is meaningful because of the Thirdness it involves. Similarly, pragmatism as a logical maxim would simply remain meaningless if it did not involve metaphysics [10] (p. 149).

The future is not given. The future may be changed via action, or via chance. Some future conditions are more favorable to me as an entity in the present than other future conditions. The choice of next actions among many possible next actions should be guided, in Peirce's view, by pragmatic considerations for three reasons. One, not all alternatives may be tested simultaneously. Two, some alternatives are more likely instrumental than others. Three, any alternative has its own unique set of actions and steps, what might be called costs. Peirce developed the pragmatic maxim to provide guidance for what we should attend to next, and how.

Peirce, the Polymath

C.S. Peirce was a man of many capabilities and many accomplishments. His contributions spanned all three of the sciences of discovery (as he named them)—mathematics, cenoscopy, and idioscopy—previously discussed. Peirce's advances in mathematics, logic, physical sciences, and the scientific method are legion. He was the first to develop a theory of signs (semiosis), is the acknowledged 'father' of American pragmatism, developed diagrammatic ways to represent logic via existential graphs, and explicated a new kind of inference, abductive reasoning. He made contributions to linguistics, the categorization of the sciences, geodesy, and topology. His precise work on physical measures with pendulums and in chemistry led him to make advances in probability, statistics, and instrument errors. He was a realist and understood the limits to truth. His advances appear grounded in a relentless questioning of premises and a rigorous application of logic to the most basic questions. These quests led him to a fundamental cosmogony built around the irreducible and universal categories of Firstness, Secondness, and Thirdness.

One of the best general introductions to Peirce's lifelong accomplishments is provided in the hard-to-find "Introductory Note" by Max Fisch to Chap. 2 of Thomas Sebeok's 1981 book, *Play of Musement* [16]. For those keenly interested in Peirce's life and accomplishments, this obscure paper is worth tracking down. One thing we do know is that Peirce was a classifier throughout his life. His classifications range among the foundations of cosmology to phenomena, relations, natural classes, sciences and knowledge, signs and knowledge representation, logic, and mathematics. In keeping with that spirit, I, too, will classify Peirce's accomplishments according to the sciences of discovery.

Mathematics

Peirce's father, Benjamin, was a noted mathematician, and Charles grew up being tutored and challenged in math, including mathematical games. Peirce made substantial contributions to the field of mathematics in many areas throughout his working career, though he did admit to backing away from rigorous mathematical problems late in his life.

Peirce's deepest and broadest contributions were in mathematical logic, where he pioneered many new areas [17]. Putnam provides a good overview of Peirce's many intellectual contributions to logic [18]. We have already noted his explication of the third mode of logical reasoning, abductive logic, and his invention of the notation and rigor of existential graphs. The term 'first-order logic' is due to Peirce. He was a keen student and critic of the leading logicians of his era, De Morgan, Boole, Schröder, and Venn. Peirce considered his work on the logic of numbers and the analysis of the infinite as being one of his major mathematical contributions [1]. Peirce also referenced a three-valued logic, many methods for which he had already developed [19]. We can also point to a second major area of contributions in probability theory, then known as the Doctrine of Chances (1878, CP 2.645-66). He explicated important ideas in randomness and sampling and analytic methods. We have already seen how important probability and continuity were to Peirce's metaphysics.

In his earlier years, Peirce developed a calculus founded on the actualness of infinity and infinitesimals. He suggested a cardinal arithmetic for infinite numbers, years before any work by Cantor (who completed his dissertation in 1867) and without access to Bolzano's 1851 work. In 1880–1881, he showed how to do Boolean algebra via a repeated and sufficient single binary operation.

In 1881 he set out the axiomatization of natural number arithmetic, a few years before Dedekind and Peano. In the same paper, Peirce gave, years before Dedekind, the first purely cardinal definition of a finite set. In that same year of 1881, Peirce provided the first successful axiom system for the natural numbers. Soon after, he distinguished between first-order and second-order quantification. In the same paper, he set out what can be read as the first (primitive) axiomatic set theory, anticipating Zermelo by about two decades. He also made contributions in the areas of finite differences and linear associative algebra [20].

Peirce was intrigued with the ideas of geometric or notational expressions of algebra, often regarding notions of continuity. He was an explicator of some of the earliest foundations of topology, and his invention of existential graphs is a direct expression of this interest. Peirce is the inventor of the quincuncial projection of the sphere. He also claimed a proof that only four colors are needed to color a spheroidal map (the so-called four-color problem).

Peirce was also the first to apply statistical methods to comparative biography [1], and he also applied his mathematical approaches to what is today known as political economy and econometrics. In 1880 Peirce was elected as a member of the London Mathematical Society.

Cenoscopy

A prior section and many references throughout this book deal with Peirce's contributions to the science of cenoscopy, or philosophy, which are legion, including his co-founding of the Metaphysical Club in Cambridge in 1872. I only want to add one further point here, and it relates to the idea of the 'highest good,' or *summum bonum*. Peirce sees striving for the *summum bonum* as moving toward the perfection of Thirdness of continuity or generality by the process of evolution:

> … the pragmaticist does not make the *summum bonum* to more consist in action, but makes it to consist in that process of evolution whereby the existent comes more and more to embody those generals which were just now said to be *destined*, which is what we strive to express in calling them *reasonable*. In its higher stages, evolution takes place more and more largely through self-control, and this gives the pragmaticist a sort of justification for making the rational purport to be general (1905, CP 5.433).

In his sciences of discovery, Peirce places esthetics and ethics with logic as the three normative sciences, with their dependence on one another:

> My own view in 1877 was crude. Even when I gave my Cambridge lectures [1898] I had not really got to the bottom of it or seen the unity of the whole thing. It was not until after that that I obtained the proof that logic must be founded on ethics, of which it is a higher development. Even then, I was for some time so stupid as not to see that ethics rests in the same manner on a foundation of esthetics… (1910, CP 8.235).

Regarding which, Peirce elaborates further on the definition of esthetics and its relations to the *summum bonum*:

> Esthetics is the science of ideals, or of that which is objectively admirable without any ulterior reason. I am not well acquainted with this science; but it ought to repose on phenomenology. Ethics, or the science of right and wrong, must appeal to Esthetics for aid in determining the *summum bonum*. It is the theory of self-controlled, or deliberate, conduct. Logic is the theory of self-controlled, or deliberate, thought; and as such, must appeal to ethics for its principles. It also depends upon phenomenology and upon mathematics (1900, CP 1.191).

Peirce alludes to what is goodness, the esthetics to which ethics impels action as the governing principle of logic, in many areas throughout his writings [21]. His ideals of looking to the community to help guide inquiry and adjudicate truth are also grounded in his practical ethics. We can see an esthetic core to the ethics that govern Peirce's overall philosophy.

Idioscopy

In the area of the special sciences, that is, the standard sciences plus the humanities (nature and mind) that Peirce termed the *idioscope*, Peirce's contributions occur in three different areas. The first area, and most prolific, was Peirce's contributions as a scientist. The second area was as an inventor. The last area of contribution comes from Peirce's special skills as a person, a humanist.

Scientist

For most of his employed life, apart from his teaching at Johns Hopkins and the piecework that constituted much of his later employment, Peirce was a practicing physical scientist. He made notable contributions in geodesy, astronomy, metrology, and chemistry. As a practicing scientist, Peirce gained much appreciation for the difficulty and lack of precision and repeatability in measurements. He understood the importance of accurate tools and measurement standards for capturing small differences. These first-hand experiences contributed greatly to his views on probabilities and the role and significance of the scientific method.

Geodesy was a primary responsibility for Peirce during his more than three-decades-long employment at the US Coast and Geodetic Survey, as introduced previously. He proposed using the wavelength of sodium light as a means to measure the length of pendulums more accurately, anticipating the metric standard by many decades. These studies also helped improve our understanding and calculation of the exact shape of the earth [17].

Peirce made many contributions to astronomy, including computations of theoretical astronomy, stellar observations, and theories of error. He was among the first to propose (correctly) that the Milky Way forms a disc, and did pioneering work on the magnitude of stars in the Milky Way [1]. The only full-length book authored solely by Peirce during his lifetime was an 181-page monograph in 1878, *Photometric Researches*, on the applications of spectrographic methods to astronomy [22].

In many areas, various researchers have noted Peirce's foresight in his scientific endeavors. For example, Ilya Prigogine claimed that Peirce's "Design and Chance" article, written in 1884, with its view of time and the second law of thermodynamics, anticipated the 'new physics' of the twentieth century. We note other areas for which Peirce foresaw or alluded to future science or discoveries throughout this book. Peirce was elected as a Fellow of the American Academy of Arts and Sciences in 1867, and a member of the National Academy of Sciences in 1877.

Inventor

Besides inventions such as map projections, semiotics, and pendulum design mentioned elsewhere, Peirce was also a prolific developer of notations and classical inventions. In notations, his existential graphs certainly stand out. However, he was also an inventor of the Peirce arrow symbol for the logical 'neither nor,' also called the Quine dagger (NOR), and its NAND complement, the Sheffer stroke. Peirce also created a unique method of iconic handwriting, which he dubbed 'Art Chirography.'

In 1892, Peirce developed an electrolytic bleaching process for wood pulp. A few years later, he also invented an acetylene lamp generator, also later tied into a hydroelectric project, that was competing with Edison's electric light. At this same time, after his dismissal from Hopkins, he also conducted stress engineering analysis for what would eventually become the George Washington Bridge in New York City [17].

His strengths in logic and his inventive mind also foreshadowed the modern computer era. Some claim that he invented the electronic switching-circuit computer [17]. In 1886, he saw that Boolean calculations could be carried out via electrical switches, anticipating Claude Shannon by more than 50 years. He also wrote on Charles Babbage and posited the use of electricity and logic gates for reasoning machines.[11]

Humanist, as Person

Along with his student Joseph Jastrow at Hopkins, Peirce was one of the first experimental psychologists in the United States, pioneering experimental studies in 'subliminal' perception. He had definite views on the concept of higher academic education as a pursuit of collective research, an approach that he embodied in the *Studies in Logic* in 1893, a collection of essays by Peirce and his students. He wrote over 300 book reviews for the *Nation* magazine, and wrote a textbook in elementary mathematics, unpublished in his lifetime, that Carolyn Eisele painstakingly recreated in the *New Elements of Mathematics* series [19, 20, 23, 24]. For many years, Peirce documented individual studies of great men (1900, CP 7.256). He proposed the *Logotheca* as an updated replacement for *Roget's Thesaurus*.

Peircean ideas have been influential in linguistics, specifically in the fields of cognitive linguistics, diachronic linguistics, linguistic semantics and pragmatics, and text linguistics, most driven by his semiotic insights [25]. Peircean ideas have also informed computational approaches to linguistics [26] and language parsing [27] (see also Chap. 16). He produced an important work on pronunciation of Shakespearean English [17]. Peirce was also an avid book collector and adviser to the New York Public Library for the purchase of scientific books [1]. As will be noted in the next section, Peirce was an accomplished lexicographer, specializing in definitions of technical topics. He also was a translator, sometimes for hire, in Greek, Latin, French, and German.

According to Brent [1], Peirce was a practiced actor, belonging to many amateur acting groups, and his wife, Juliette, was reportedly an actress of some ability. He was lauded at times as a storyteller, orator and debater, teacher, and lecturer, though

[11] References to Charles Babbage may be found at CP 2.56 and CP 4.611. For electrical logical machines, see Charles S. Peirce, 1993, "Letter, Peirce to A. Marquand" dated 30 December 1886, in Kloesel, C. et al., eds., Writings of Charles S. Peirce: A Chronological Edition: Volume 5: 1884–1886. Indiana University Press: 421–422, with an image of the letter page with the circuits on p. 423.

other occasional reports characterize certain of his lectures as rambling, unintelligible, or dislocated. Peirce even knew card tricks and practiced occasional magic tricks. He was very much interested in mazes and games and published a series in *The Monist* on "Amazing Mazes" later in his life. As a hobby and because of family illnesses, Peirce was also well versed in the history and theory of medicine.

An Obsession with Terminology

Though Peirce frequently railed against nominalism, arguing instead for a realistic view of the world, he was also very attuned to names, labels, and definitions. He sought the 'correct' way to label his constructs. As one instance, at various times, Peirce called abductive reasoning *hypothesis*, *abduction*, *presumption*, and *retroduction.* He also called the methodeutic *speculative rhetoric*, *general rhetoric*, *formal rhetoric*, and *objective logic*. Such changing names were not uncommon with Peirce.

In his lifetime, Peirce both enjoyed and made money as a lexicographer defining terms. He personally wrote 6000 entries for the 12-volume *Century Dictionary* [28], and oversaw a total of 16,000 entries where he had primary responsibility in such areas as logic, mathematics, mechanics, measurement, philosophy, astronomy, and universities [17]. Peirce held that the understanding of a language symbol is a process of shared consensus among its community of users; he loathed to use common terms for his constructs. Indeed, when one of his terms, pragmatism, was adopted by William James who gave it a different spin and interpretation, Peirce disavowed his earlier term and replaced it with the term *pragmaticism*.

> So then, the writer [Peirce], finds his bantling 'pragmatism' so promoted, feels that it is the time to kiss his child good-bye and relinquish it to its higher destiny; while to serve the precise purpose of expressing the original definition, he begs to announce the birth of the word 'pragmaticism', which is ugly enough to be safe from kidnappers (pp 165–166) [29].

Peirce should have realized that understandability holds sway over individualized perspective. He was silly to argue with James about the term pragmatism, as James was doing so much to promote awareness of Peirce's ideas.

Table A.1 Examples of obscure Peirce terminology

Agapism	Cenoscopy	Interpretant	Phaneroscopy	Semeiotic
Anancasticism	Cyclosy	Legisign	Pragmastic Definition	Sinsign
Apeiry	Dicent	Medisense	Pragmaticism	Speculative rhetoric
Antethics	Entelechy	Methodeutic	Precission	Stecheotic
Architectonic	Fallibilism	Objective Idealism	Qualisign	Synechism
Axiagastics	Hylozoism	Percipuum	Representamen	Transuasion
Ceno-pythagorean	Hypostatic Abstraction	Periphraxy	Retroduction	Tychasticism
Chorisy	Idioscopy	Phaneron	Rheme	Tychism

This penchant for 'ugly' terms was not uncommon for Peirce. As examples, Table A.1 presents some terminologies from Peirce's writings. Changing and 'ugly' terminology is but the first of the difficulties in reading and understanding Peirce. His evolution as a thinker, plus the interpretations of those who study Peirce, also complicates matters. A real point about interpretation, I think, is to try to get past his sometimes off-putting terminology. Mostly what is hard to understand are terms you may be encountering for the first time.

I can appreciate Peirce's preference for precision in how he described things. I can also understand scholars sometimes concentrating more on literalness than meaning. But the use of obfuscatory terms or concentrating on labels over the conceptual is a mistake. When looking for a precise expression of new ideas I try to harken to key Peircean terms and concepts, but I sometimes find that alternative descriptions within Peirce's writings better communicate to modern sensibilities. Concepts attempt to embody ideas, and while it is useful to express those concepts with clear, precise and correct terminology it is the idea that is real, not the label. In Peirce's worldview, the label is only an index. I concur. In the semantic Web, we sometimes refer to this as 'things, not strings.'

Peirce, the Polestar

That we live in an age of information and new technologies and new developments is a truth evident to all. These developments lead to a constant barrage of new information, often leading to new or revised assertions ('facts'). What we believe and how we interpret that information are what we call knowledge. New facts connect to or change our understanding of old 'facts'; those connections, too, are a source of new knowledge. Our (1) powers of observation and learning and discovery; (2) interactions and consensus-building with communities; and (3) the methods of scientific inquiry, all cause us to test, refine, and sometimes revise or discard what we thought were prior truths. Knowledge is thus dynamic, continually growing, and subject to revision.

What I call a Peircean *mindset* can help inform answers to new problems, problems that Peirce did not directly address himself. Indeed, the problems that set this context are machine learning and natural language understanding, all driven by computers and electronic data unimagined in Peirce's day. Because my views come from my own context, something that Peirce held as an essence of Thirdness, I cannot say that I *base* my views on Peirce's views. Who knows if he would endorse my views more than a century after his death? However, my take on these matters is the result of much reading, thought, repeat reading, and study of Peirce's writings. So while I cannot say I *base* my views on Peirce, I can certainly say that he *guides* me.

Peirce's universal categories of Firstness, Secondness, and Thirdness provide the mindset for how to think about and organize knowledge. The tasks of defining and organizing knowledge demand that we bring meaning, context, and perspective to the task. I believe Peirce's universal categories and what they imply offer the next adaptive climb upward for knowledge representation. The overarching framework of Peirce's philosophy—his architectonic—is grounded in these categories. As a scientist and logician, Peirce applied this mindset in pragmatic and testable ways. These methods, indeed the scientific method itself, further guide how and where to apply this mindset in ways that are economical and promise the most knowledge among all of the possible paths of inquiry.

Peirce's fierce realism, his belief in reality beyond our minds, and his insistence that this reality is subject to inquiry and the fixation of belief leading ever closer to truth are distinctly different from the mind-body duality put forward by Descartes. Richard Bernstein in a recent book [30] called this viewpoint a sea change:

> Pragmatism begins with a radical critique of Cartesianism. In one fell swoop, Peirce seeks to demolish the inter-related motifs that constitute Cartesianism [mind-body duality; primacy of personal experience; doubt as a starting condition; there are incontrovertible truths to be discovered] … We can view the development of pragmatism from Peirce until its recent resurgence as developing and refining this fundamental change of philosophical orientation—this sea change. A unifying theme in all the classical pragmatists as well as their successors is the development of a philosophical orientation that replaced Cartesianism (in all its varieties) (pp 18–19).

Our real world is always changing, continually unfolding. Our real world is viewed by all of us differently, based on background, predilection, perspective, and context. What we think we know about the world today is subject to inquiry and new insights. New factors are arising to shift what we think we know about ourselves and our place in the world.

How I interpret Peirce, and why he has become a polestar in my thinking about knowledge representation, embraces three perspectives. First, given the breadth of Peirce's insights, I try to read as much by him and about his writings by others as I can. This exposure helps set a fertile milieu, useful to interpretation and critical judgment. Second, despite my awe of Peirce's genius, I do not treat his writings as gospel. Were he alive today, I do not doubt that the massive increase in knowledge and information since his day would cause him to alter his viewpoints—perhaps substantially so in some areas. Third, no similar reason compels us to shy away from questioning any of Peirce's assertions. Nonetheless, given Peirce's immense powers of logic, one better be well prepared with evidence and sound reasoning before undertaking such a challenge.

Resources About Peirce

Slowly at first, and then growing after the publication of the *Collected Papers*,[12] a legion of researchers and academics have labored to preserve, understand, and explicate Peirce's insights. Virtually every author and name mentioned in this book have played such a role, with hundreds more, some more active than those cited, contributing their part. I share in this section some of my preferences and personal selections for useful resources about Peirce.

My first recommendation to begin learning about Peirce is to start with Wikipedia. Its English entry on Charles Sanders Peirce is quite good and rather complete. An entire category is dedicated to Peirce on Wikipedia, with some 40 articles listed. I think the articles on semiosis, abductive reasoning, and pragmaticism are some of the better ones. Unfortunately, the article on Peirce's universal categories is pretty weak. To compensate, however, the Wikipedia Peirce bibliography is an excellent reference source.

Peirce is hardly easy to read, and most of what others write about him is also pretty dense. Though those seasoned in Peirce studies might find it covering standard ground, the 2013 Cornelis de Waal guide to Peirce [17] is an accessible introduction to Peirce and his contributions. I no longer consult it for facts or details, but as an intro it is helpful and a relatively quick read. If this piques your interest, then it is probably worth your time to start exploring Peirce in more depth. I also like de Waal's labeling of the 'doctrine of the categories.'

After introductions, it is best to study Peirce in his own words. The earliest known compilation of Peirce's writings was by Cohen in 1923 [31], nearly a decade after Peirce's death, and is both a good intro and starting compilation. An even better starting compilation is that of Buchler [32]. However, I did not start with either of these nor with de Waal, because my initial research discovered that searchable PDF versions of the first 'complete' compilations of Peirce's writings could be obtained for free online [33]. The *Collected Papers* are available online in a version easier to read than the PDF versions, and which you can search.[13] The problem with these CP sources, however, is that the editorial order of CP is not chronological, gaps exist because of the sources initially chosen, and the formatting and editorial decisions are not equal to later standards. The online version is better for learning and reading purposes, but the lack of editorial oversight hurts CP irrespective of format. (A prior CD library is also no longer available [28].)

[12] The jumbled nature of the original *Collected Papers* means they should be used with caution, since they have no chronology. Many contemporary Peirce scholars now tend to date by year the passages they quote in order to overcome this problem.

[13] You can use Google to search within the textlog.de site, even though it is in German and does not have its own search function, by using a query similar to the following: https://www.google.com/search?hl=en&as_q=peirce+abduction&as_sitesearch=www.textlog.de. Note: Include "peirce" in the request, because there are other philosopher papers on the textlog.de site. Also note that this approach is tailored for English, with the example querying for "abduction"; replace your own search query in the query string.

Of course, many editors have compiled Peirce's writings. In mathematics, you likely want to focus on the fantastic four-volume series from Eisele [34], which can often be found for free online. As a nonmathematician, I found Volume 4 the most useful. For my interests in logic and knowledge representation, I have found *Vol 1* of *The Essential Peirce* [35] the best single compilation of relevant writings. In fact, you can reassemble the entire contents of EP (as it is abbreviated) from free, online PDFs, and I have, but that also means you lose the fantastic Nathan Houser introduction and the excellent packaging and portability of an actual paperback book. Many other compilations are also available (see the various bibliographic sources).

I almost uniformly find the introductions by the editors of these compilations provide useful insights about Peirce. The introductions often weave in relevant personal details to help evaluate Peirce as a person. The editors bring a perspective and context to Peirce's accomplishments since they offer an external vantage. Under the category of editorial compilations, I especially like Nathan Houser's introduction to EP. However, from different perspectives, the intros by both Brent and Murphey (see below) helped bring him alive to me.

After this kind of a dive into Peirce's writings, again usefully supported by the editor's intros, I find I want a big picture of Peirce, which covers his motivations, circumstances, discoveries, and maturation. I suspect these are the hardest of the books about Peirce to write. It requires a breadth of familiarity and a deep understanding of (at least what the author thinks are) Peirce's intentions. There also are variants of this approach, focusing on specific slices (such as religion [36]) or particular concepts or academic perspectives.

The online Arisbe, the Peirce Gateway,[14] lists some 210 books published on Peirce and related topics since 1995 or so, with 114 published since 2006 alone. The site further lists 357 doctoral dissertations about Peirce, most in the last few decades. Note that many of these sources are not in English since Peirce is studied worldwide, with a strong following in Latin America, especially Brazil and Colombia. The Arisbe site is helpful in that most entries include at least a paragraph of description, and often with links to more extended online excerpts. Arisbe is a good resource should specific topics pique your interest while studying Peirce.

Among the comprehensive studies covering the entirety of Peirce's life work, I will mention two. The first is the book from Kelly Parker in 1998 [7] that focuses on Peirce's emphasis on continuity (synechism). Parker writes well, is lucid, and has an excellent notes section. The second compilation, and one of my favorite Peirce reads, is the earlier 1993 book by Murphey [6] on the development of Peirce's philosophy. Some other scholars, notably Hillary Putnam, have suggested that Murphey's interpretations are often controversial. Murphey did, indeed, change some of his opinions of Peirce, especially about continuity, in the second edition. However, I find Murphey's analysis of the phases of Peirce's developments to conform to my sense. The latter section of his book is excellent. I find it strange that many other general recommendations for Peirce readings tend to overlook this

[14] http://www.iupui.edu/~arisbe/

book. Perhaps a bit of this neglect came from Putnam's early comments, but Murphey is one of the resources I most often consult.

When first learning about Peirce, it is striking how dominant semiosis and his theory of signs (and logic) pervade many of the resources. These are critical Peircean topics, but I find that it took me a while to probe beyond these topics into others I find even more fascinating. I have focused on Peirce's universal categories of Firstness, Secondness, and Thirdness. I have also been studying abductive reasoning, language grammars, link between logic and mathematics, and how Peirce's views dovetail into current topics in topology and category theory. About these last topics, I recommend Fernando Zalamea [37]. Zalamea's scholarship is quite advanced, and I do not recommend as a starting point, but after some exposure to Peirce I like the synthetic view that Zalamea brings to the table. His scholarship shows that Peirce continues to produce major insights for modern logic and mathematics.

Biographies are another useful source. Louis Menand won a Pulitzer Prize for his recounting of the birth of pragmatism in the United States [38]. He told the story through the lens of the major participants in the Metaphysical Club, really more of an informal grouping of intellectuals. William James, Chauncey Wright, and Oliver Wendall Holmes figured prominently in that group, but none perhaps more so than Peirce. (Peirce and James were lifelong friends, but Peirce tremendously respected Wright for his insight and intellect, and they were very close friends; Wright, unfortunately, died young at 45.) What is great about this book is that the author frames the movement to pragmatism through the prism of slavery and abolition, the Civil War, and rapid intellectual and technological change. This perspective makes for an excellent read because it does such a marvelous job of placing Peirce into the context of his times, as well as providing equivalently fascinating looks at his very accomplished colleagues. However, this is not the single book to read if you want to probe deeply into Peirce's theories and worldview.

My favorite biography of Peirce, whose publication is a pretty astonishing story in its own right, is Brent's life biography of C.S. Peirce [1]. Brent first began his biography of Peirce to answer the question of who invented the US philosophy of pragmatism, triggered by clues in a biography of Peirce's friend, William James. He completed his dissertation in 1960 and intended to publish it, but ran into permission difficulties from Harvard, which was still acting poorly about Peirce's archival papers. Brent had to drop the project and moved on to other things. Then, in 1988, Thomas Sebeok, himself a then emerging Peirce scholar, encountered a description of the dissertation in a footnote in another book. He was able to get the dissertation through interlibrary loan and finally read it in 1990. He was astonished by what he learned and the quality of the work and set out to find Brent, whom he eventually tracked down in Washington, DC. Through Sebeok's catalyst, a publisher was found; Brent agreed to update his 30-year-old dissertation, itself an effort of considerable labor, and the work was finally published in 1993. Brent provides an unvarnished and, I believe, fair look at Peirce, the person, and shows insight into his accomplishments and unique ways of thinking about the world. Brent tackled head on all of Peirce's foibles and weaknesses as well. The resulting biography is a masterpiece, what Sebeok termed a 'tragicomic thriller.' Brent himself came to believe "in philosophy [Peirce] was one of the most original thinkers and system builders of any time, and certainly the greatest philosopher the United States has ever seen." Brent came to feel 'deep affection' for his subject, despite those foibles and weaknesses. However, some Peirce scholars, such as Gary Richmond, think the biography unfair, with too much prominence given to Peirce's critics.

The Brent biography is an incredibly intelligent treatment of an incredibly intelligent man. As might be expected from a work that began as a dissertation, it is thorough and well referenced. As might not be expected from a dissertation, it is well written. Brent uses Peirce's own 'architectonic,' a term new to me then but studied by me now, a term drawn from Aristotle but modified by Kant and then Peirce, as a way of framing his treatment. Brent is also attuned to shifts in Peirce's thinking over time, a great boon to better understand the development of his theories. Since I believe others will study Peirce for centuries, along with other great thinkers of humankind, Brent's biography will be a must-include companion to Peirce's writings. As I note in the close to this chapter, Brent and Sebeok are but two of the hundreds of individuals who have made it their life's work and passion to understand Peirce better, to convey what he was trying to tell us, and to bring awareness of him to broader audiences. Also, a fictionalized biography of Peirce's mysterious second wife, Juliette, has some voyeuristic interest but is an unsuitable source for any reliable information about either Charles or Juliette [39].

The bulk of commentary, of course, about Peirce may be found in the academic literature. I often find when studying Peirce that a new topic (or one that finally gets my attention) will arise about which I want to learn more. As with all such topics, I first consult Wikipedia for a starting article, if one exists, to get a bit of background and then some key links and useful search terminology. However, my real focus in such investigations centers on Google Scholar. Google Scholar contains nearly 40,000 articles about or discussing Peirce, with the bulk, perhaps 70%, in English.[15] When searching Scholar, I always use "peirce" as one of my keywords and keep that search term in quotes (without the quotes, Scholar will also give you results from "pierce" since it seems to assume "peirce" is a misspelling). For papers of keen importance, I will also click the link 'Cited by xx' link on Scholar and do a secondary inspection of those to find other interesting papers that have cited the one of interest. I have assembled a complete electronic Peirce library of hundreds of documents over time in this manner.

A society is devoted to Peirce. One may find many Web sites such as Arisbe at the University of Indiana, online forums including for biosemiotics, annual conferences, and many individuals with their Web sites and writings who analyze and

[15] Here is an example query: https://scholar.google.com/scholar?q="peirce"+abduction. Substitute your own topic keywords for "abduction" in the example query string.

pronounce strong views as to what Peirce meant and how we should interpret him. I have often mentioned the influence of John Sowa in first getting me interested in Peirce, so his site (with query specific to Peirce) is a good one to include on your list. Sowa tends to focus on existential graphs, knowledge representation, logic, and natural language understanding. You may also find a good source for Web writers on Peirce on the Arisbe site; check out the blogroll on the left column. Of course, I, too, write not infrequently about Peirce. You may obtain my Peirce articles under my blog's Peirce category. I hope the dozen or so others who often write on Peirce forgive me for not directly mentioning them. Thank you, and I hope we see more.

For broad electronic resources on Peirce, probably the best is Arisbe, noted already.[16] Two high-quality, online philosophy sites, the Stanford Encyclopedia of Philosophy and the Internet Encyclopedia of Philosophy, are often useful introductory resources when beginning to learn about a new topic. Authoritative scholars write many of the Peirce articles. A site not updated since the early 2000s, but which has some unique and high-quality articles by outside experts, is the Digital Encyclopedia of Peirce. A useful site to see some different uses of specific Peircean terms is the Commens Web site. The Charles S. Peirce Project was established in 1976 to continue the mission of making Peirce's writings available, started by the *Collected Papers* (CP) project [33] going back to the 1930s. The Project continues to produce a multivolume chronological and critical edition of Peirce's writings. Romanini's Minute Semeiotic Web site is a fun way to explore what Peirce might have intended with his (incomplete) 66-sign schema.

Since first established by Joe Ransdall in 1993, a dedicated discussion list, Peirce-L, with often lively discussion, has nearly daily activity. That link will allow you to search archives going back to 2011, and to subscribe to the list. A similar mail list exists for a group in biosemiosis, another field that Peirce played no small role in helping to gestate. A useful piece of information if you study Peirce further, given that so much of his writing appeared long ago or has been transcribed or compiled by editors, is how to decipher the citation schemes. Good sources on Peirce citation standards are the Wikipedia CSP abbreviations and Robin catalog for citing papers and manuscripts. For the truly dedicated, you can help crowd-translate Peirce's unpublished manuscripts via the SPIN project co-directed by the Peirce scholar, Jeffrey Brian Downard.

[16] See http://www.iupui.edu/~arisbe/faqs/whyarisb.HTM for the history of the term Arisbe as used by Peirce for his Pennsylvania home.

References

1. J. Brent, *Charles Sanders Peirce: A Life* (Indiana University Press, Bloomington, 1998)
2. N. Houser, The fortunes and misfortunes of the Peirce papers, in *Signs of Humanity*, ed. by M. Balat, J. Deledalle-Rhodes (Mouton de Gruyter, Berlin, 1992)
3. C.S. Peirce, Peirce edition project. The ten main trichotomies of signs, in *The Essential Peirce – Volume 2: Selected Philosophical Writings (1893–1913)* (Indiana University Press, Bloomington)
4. S. Case, The man with a kink in his brain, in *National Review (online)* (2014), vol. 21
5. M.H. Fisch, Peirce's Arisbe: the greek influence in his later philosophy, in *Peirce, Semiotic, and Pragmatism: Essays*, ed. by M. H. Fisch, K. L. Ketner, C. J. Kloesel (Indiana University Press, Bloomington, 1986), p. 227
6. M.G. Murphey, *The Development of Peirce's Philosophy* (Hackett Publishing Company, Inc., Indianapolis, 1993)
7. K.A. Parker, *The Continuity of Peirce's Thought* (Vanderbilt University Press, Nashville, 1998)
8. C.S. Peirce, On a new list of categories, in *Proceedings of the American Academy of Arts and Sciences* (1867)
9. B.E. Kent, *Charles S. Peirce: Logic and the Classification of the Sciences* (McGill-Queen's University Press, Kingston, 1987)
10. S. Ika, *A Critical Examination of the Philosophy of Charles S. Peirce: A Defence of the Claim that his Pragmatism is Founded on his Theory of Categories*. Ph.D., University of Notre Dame Australia (2002)
11. P. Rose, Another guess at the riddle: more ado about nothing, in *Analecta Hermeneutica* (2013)
12. C.S. Peirce, The order of nature. Pop. Sci. Monthly **13**, 203–217 (1878)
13. S. Haack, Not cynicism, but synechism: lessons from classical pragmatism. Trans. Charles S. Peirce Soc. **41**, 239–253 (2005)
14. J.L. Esposito, Synechism: the keystone of Peirce's metaphysics, in *Digital Encyclopedia of Charles S. Peirce* (2001)
15. M.H. Fisch, Peirce's Progress from Nominalism Toward Realism. Monist **51**, 159–178 (1967)
16. M.H. Fisch, Introductory note, in *The Play of Musement* (Indiana University Press, Bloomington, IN, 1981), pp. 17–21
17. C. De Waal, *Peirce: A Guide for the Perplexed* (A&C Black, 2013)
18. H. Putnam, Peirce the Logician. Hist. Math. **9**, 290–301 (1982)
19. C.S. Peirce, *The New Elements of Mathematics – Volume 3* (Mouton, The Hague, 1976)
20. C.S. Peirce, *The New Elements of Mathematics – Volume 4* (Mouton, The Hague, 1976)
21. J. Feibleman, A Systematic Presentation of Peirce's Ethics. Ethics **53**, 98–109 (1943)
22. M. Bergman, C. S. Peirce: a short biographical sketch, in *Commens: Digital Companion to C.S. Peirce* (undated)
23. C.S. Peirce, *The New Elements of Mathematics – Volume 1* (Mouton, The Hague, 1976)
24. C.S. Peirce, *The New Elements of Mathematics – Volume 2* (Mouton, The Hague, 1976)
25. W. Nöth, Charles sanders Peirce, pathfinder in linguistics, in *Digital Encyclopedia of Charles S. Peirce* (2000)
26. G. Richmond, Outline of trikonic diagrammatic trichotomic, in *13th International Conference on Conceptual Structures, ICCS 2005*, ed. by F. Dau, M. L. Mugnier, G. Tumme (Springer-Verlag GmbH, Kassel, 2005), pp. 453–466
27. J.F. Sowa, Toward the expressive power of natural language, in *Principles of Semantic Networks* (Elsevier, Oxford, 1991), pp. 157–189
28. C.S. Peirce, *The Writings of Charles S. Peirce—A Chronological Edition*. Electronic Edition, InteLex
29. C.S. Peirce, in *The Monist.* What Pragmatism Is (1905), pp. 161–181
30. R.J. Bernstein, *The Pragmatic Turn*, Polity (2010)

31. C.S. Peirce, K.L. Ketner, *Chance, Love, and Logic: Philosophical Essays* (Lincoln, University of Nebraska Press, 1998)
32. C.S. Peirce, *Philosophical Writings of Peirce: Selected Writings* (Routledge and Kegan Paul Ltd., reissued by Dover Publications, New York, NY, 1940)
33. C.S. Peirce, *The Collected Papers of Charles Sanders Peirce*, vol 8 (Harvard University Press, Cambridge, MA, 1931)
34. C.S. Peirce, *The New Elements of Mathematics by Charles S. Peirce* (The Mouton Publishers, The Hague, 1976)
35. C.S. Peirce, *The Essential Peirce: Selected Philosophical Writings, Vol 1 (1867–1893)* (Indiana University Press, Bloomington, 1992)
36. M. Raposa, *Peirce's Philosophy of Religion* (Indiana University Press, Bloomington, 1993)
37. F. Zalamea, *Peirce's Logic of Continuity: A Conceptual and Mathematical Approach* (Docent Press, Boston, 2012)
38. L. Menand, *The Metaphysical Club: A Story of Ideas in America* (New York, NY, Farrar, Straus and Giroux, 2001)
39. M. Samuels, *The Queen of Cups* (Unlimited Publishing LLC, Bloomington, 2006)

Appendix B: The KBpedia Resource

KBpedia is a computable knowledge structure resulting from the combined mapping of six, large-scale, public knowledge bases—Wikipedia, Wikidata, OpenCyc, GeoNames, DBpedia, and UMBEL.[1] The KBpedia structure separately captures entities, attributes, relations, and concepts. KBpedia classes these into a natural and rich diversity of types, with their meaning and relationships, logically and coherently organized into about 80 typologies.

KBpedia is the first full-blown ontology based on Charles Sanders Peirce's universal categories and logic of relations. The KBpedia knowledge structure is written in the OWL semantic language; all underlying structures are represented in either OWL or RDF. KBpedia follows best practices, many of which were pioneered by KBpedia's editors, governing knowledge, and concept representation and annotations. All languages and knowledge representations are written in W3C-compliant standards.

The focal objective of KBpedia is to exploit large, public knowledge bases to support artificial intelligence using both supervised and unsupervised machine learning methods. KBpedia is explicitly designed to expose rich and meaningful feature sets to support the broadest range of machine learning methods. KBpedia is also specifically structured to enable useful splits across a myriad of dimensions from entities to relations to types that can all be selected to create positive and negative training sets, across multiple perspectives. The disjointedness of the SuperTypes that organize the 55,000 entity types in KBpedia provides a powerful selection and testing mechanism.[2] The coherency of KBpedia provides a basis for logic tests to further improve accuracy, including the creation of local gold standards, at an acceptable cost.

[1] Material in this appendix is drawn with permission from the KBpedia Web site at http://kbpedia.org

[2] This number is as of version 1.60 of KBpedia, based on the completion of this book in the first half of 2018. The current publicly released version of KBpedia likely has a different number of reference concepts. Back-of-the-envelope estimates suggest that KBpedia should eventually grow to be on the order of 80,000 reference types.

M. K. Bergman, *A Knowledge Representation Practionary*,
https://doi.org/10.1007/978-3-319-98092-8

KBpedia is a continually evolving reference structure for knowledge representation and management. KBpedia is staged to provide working levels of interoperability for the linked data ecosystem. Artificial intelligence (AI) and machine learning are revolutionizing knowledge systems. The most important factor in knowledge-based AI's renaissance has been the availability of massive digital datasets for the training of machine learners. Wikipedia and data from search engines are central to recent breakthroughs. Wikipedia is at the heart of Siri, Cortana, the former Freebase, DBpedia, Google's Knowledge Graph, and IBM's Watson, to name just a prominent few AI question answering or virtual assistant systems. Natural language understanding is showing impressive gains across a range of applications. To date, all of these examples have been the result of bespoke efforts leveraging Wikipedia, in whole or part. The tens of millions of instances captured by Wikidata add an entire ABox component to the knowledge structure. It is costly for standard enterprises to leverage these knowledge resources on their own.

Today's practices for leveraging these resources pose significant up-front and testing effort. Much latent knowledge remains unexpressed and not readily available to learners; it must be exposed, cleaned, and vetted. We need to spend further up front effort on selecting the features (variables) used and then to label the positive and negative training sets accurately. Without 'gold standards'—at still more cost—it is difficult to tune and refine the learners. The cost to develop tailored extractors, taggers, categorizers, and natural language processors is too high. KBpedia is meant to systematize a starter reference structure that new users may tailor to local domains at lower costs. Users may then apply integration and interoperability to structured, semi-structured, and unstructured data, that is, everything from text to databases. KBpedia proves that existing knowledge bases can be staged to automate much of the tedium and reduce the costs now required to set up and train machine learners for knowledge purposes. Besides labeled training sets for supervised machine learning, KBpedia, with its rich feature sets across all aspects of the knowledge structure, is also an excellent basis for selecting training corpora for unsupervised learning. It is often advisable to include some initial unsupervised learning in a more general supervised learning context.

Components

KBpedia is organized into a knowledge graph, KKO, the KBpedia Knowledge Ontology, with an upper structure based on Peircean logic. KKO sets the umbrella structure for how KBpedia's six constituent knowledge bases are related to the system. One of the three major branches of KKO, the Generals, represents the types in the system, with about 85% of the KBpedia's reference concepts residing there. These RCs are themselves entity types—that is, 47,000 natural classes of similar entities such as 'astronauts' or 'breakfast cereals'—which we organize into about

30 ‘core’ typologies (among the 80 or so) that are mostly disjoint (nonoverlapping) with one another. The typologies provide a flexible means for slicing and dicing the knowledge structure; the entity types provide the tie-in points to KBpedia’s millions of individual instances. The separate *Glossary* defines many of the terms used by KBpedia; Chap. 8 provides a more detailed discussion of KBpedia’s vocabulary.

The KBpedia Knowledge Ontology (KKO)

We inform the upper structure that is the KBpedia Knowledge Ontology (KKO) using the triadic logic and universal categories of Charles Sanders Peirce. This trichotomy, also the basis for his views on semiosis (or the nature of signs), was in Peirce’s view the most primitive or reduced manner by which to understand and categorize things, concepts, and ideas. We devote Chap. 6 to the universal categories and touch upon them and semiosis throughout the main book. We express KBpedia’s knowledge base grammar in the semantic Web language of OWL 2. Thus, we may apply most W3C standards to the KBpedia structure. The resulting, combined structure, as shown in Fig. B.1, brings consistency across all source knowledge bases.

This diagram, drawn from KBpedia’s online demo, shows the main topic areas, or typologies, that tie into KKO, which we list in their entirety in the *Structure* section below. The structure of KKO makes it a computable knowledge graph that supports inference, reasoning, aggregations, restrictions, intersections, and other logical operations. KKO’s logic basis provides a powerful way to represent individual things, classes of things, and how those things may combine or emerge as new knowledge [1].

The KBpedia Knowledge Bases

Six, large-scale public knowledge bases are central to the KBpedia knowledge structure. These six sources are as follows:

- Wikipedia—five million articles that capture the fundamental concepts and entities of basic human knowledge, often including structured data and with many linkages
- Wikidata—structured data records for about 40 million individual entities
- OpenCyc—an extract of Cyc that represents the common sense and vetted relationships among KBpedia’s base 55,000 concepts
- DBpedia—a machine-readable version of parts of Wikipedia in RDF
- GeoNames—a geographical database of some ten million places linked to about 800 distinct feature classes
- UMBEL—the initial organizational structure for the knowledge graph.

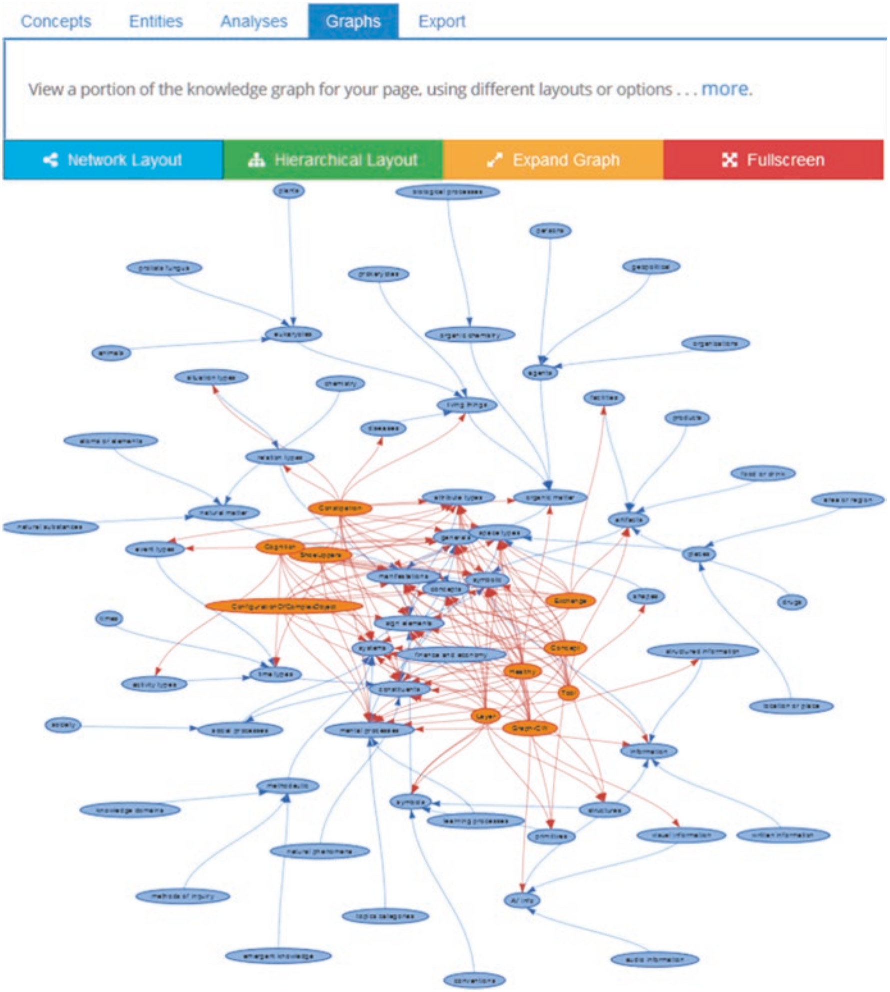

Fig. B.1 KBpedia (KKO) Network Graph

We have mapped and re-expressed each of these sources into the single, coherent knowledge system of KBpedia. We split the resulting KBpedia knowledge graph along the lines of concepts and topics, entities, events, attributes, annotations, and relations and their associated natural classifications or types. This resulting combination gives KBpedia a rich set of structural components.

Wikipedia and Wikidata are the two most important KB sources, Wikipedia for concepts, Wikidata for instances and properties, and both for multilingual capabilities. Certain aspects of Wikipedia have proven their usefulness for general knowledge acquisition, for example, using article (concept or entity) content to inform topical tagging using explicit semantic analysis (ESA) [2], automatic topic identification [3], information extraction [4], or a myriad of others. A weakness of Wikipedia has been its category structure, which was not part of the original design but added in 2004. Various reviewers have likened Wikipedia more to a thesaurus than a classification scheme, others that it is different from classical knowledge organization systems in that it has no specified root or hierarchy. This situation improved a wee bit from 2006 to 2010, when editors organized the main Wikipedia topics according to top-level and main topics.[3] Still, typical commentaries point to the fact that Wikipedia's category structure is "noisy, ill-formed, and difficult to make sense of" [4]. Its crowdsourced nature has led to various direct and indirect cycles in portions of the category structure. All of these problems lead to the inability to do traditional reasoning or inference over the Wikipedia category graph [5]. We have done much to clean the Wikipedia categories and remap them and their instances to KBpedia, which have now made the structure computable.

The choice of Wikipedia's founders to make its full content available electronically for free and without restriction was a masterstroke, and now carries over to Wikidata, a sister project to Wikipedia under the Wikimedia banner. I only hope we can honor this philosophy. Wikidata takes as its starting point the structured data about entities evident in Wikipedia infoboxes. Rather than extracting and cleaning that entity information as DBpedia does, the roles of Wikidata are as a stand-alone reference base and as a multilingual source for all entities feeding the Wikimedia network, including Wikipedia. As of June 2018, Wikidata contained about 50 million data items in its system. The Wikidata approach leads to more uniformity and consistency and provides a central Wikimedia access point for structured data [6]. However, somewhat akin to Wikipedia, Wikidata also has struggled to find an appropriate typology (or ontology) for its millions of entities.[4] Again, KBpedia provides one such structure.

Besides these main six KBs, KBpedia has extended mappings to a further 20 other vocabularies, including schema.org, Dublin Core, the Bibliographic Ontology (BIBO), and others. KBpedia also supports exports in various formats and finite-state transducers or specialty lists, used as inputs to third-party analysis, management, and visualization tools. We also transform external and domain data into KBpedia's internal canonical forms for interacting with the overall structure.

The KBpedia Typologies

A *typology* is a grouping of similar types, sharing some essential characters. Each type is a parent to a particular group of instances, which also share essential traits or attributes. Chapter 10 is devoted to a discussion of KBpedia's typologies and the advantages of their modular, expandable design. *Table 10.2* provides a listing of the current typologies; *Table 10.3* describes those that are core.

[3] For Wikipedia's main topics, see http://en.wikipedia.org/wiki/Category:Main_topic_classificationsReference; for Wikipedia's top-level categories, see http://en.wikipedia.org/wiki/Category:Fundamental_categories

[4] As of May 2018, Wikidata contained more than 54 million entities. However, there has yet to emerge an overarching typology or ontology for these entities, with the typing system that does exist growing from the bottom up. For some background, see https://www.wikidata.org/wiki/Wikidata:Requests_for_comment/Migrating_away_from_GND_main_type

Structure

This section goes into a bit more detail on the structure of the KBpedia knowledge graph, KKO. At each level in the KBpedia Knowledge Ontology, we strive to organize each entry according to Peirce's universal categories of Firstness (1ns), Secondness (2ns), and Thirdness (3ns). Most of the reference concepts (RCs) in KKO are organized under the Generals (3ns) branch, though that is not evident from inspection of the upper nodes alone. All of KKO's SuperTypes (typologies) reside there.

Most of the 30 or so core typologies in KBpedia do not overlap with one another, what is known as disjoint. Disjointness enables us to perform powerful reasoning and subset selection (filtering) on the KKO graph. Upper typologies are useful to organize core entities, plus they provide homes for shared concepts. Living Things, for example, can capture concepts shared by all plants and animals, by all life, which then enables better segregation of those life forms. We apply such natural segregations across the KKO structure. Here is the upper structure of KKO with its 171 concepts (Table B.1):

Table B.1 The KKO upper structure organized by the universal categories

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
Monads [1ns]						
	FirstMonads [1ns]					
		Suchness [1ns]				
			Accidental [1ns]			
			Inherent [2ns]			
			Relational [3ns]			
		Thisness [2ns]				
			Chance [1ns]			
			Being [2ns]			
			Form [3ns]			
		Pluralness [3ns]				
			Absolute [1ns]			
				Inclusive [1ns]		
				Exclusive [2ns]		
				Difference [3ns]		
			SimpleRelative [2ns]			
			Conjugative [3ns]			
	DyadicMonads [2ns]					
		Attributives [1ns]				
			Oneness [1ns]			
				Identity [1ns]		
				Real [2ns]		
					Matter [1ns]	
					SubstantialForm [2ns]	
					AccidentalForm [3ns]	

(continued)

Table B.1 (continued)

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
				Fictional [3ns]		
			Otherness [2ns]			
			Inherence [3ns]			
				Quality [1ns]		
				Negation [2ns]		
				Intrinsic [3ns]		
		Relatives [2ns]				
			Concurrents [1ns]			
			Opponents [2ns]			
			Conjunctives [3ns]			
				Quantity [1ns]		
					Values [1ns]	
						Numbers [1ns]
						Multitudes [2ns]
						Magnitudes [3ns]
					Discrete [2ns]	
					Complex [3ns]	
				Subsumption [2ns]		
				Connective [3ns]		
					Unary [1ns]	
					Binary [2ns]	
					Conditional [3ns]	
		Indicatives [3ns]				
			Iconic [1ns]			
			Indexical [2ns]			
			Associative [3ns]			
				Denotative [1ns]		
				Similarity [2ns]		
				Contiguity [3ns]		
	TriadicMonads [3ns]					
		Representation [1ns]				
			Icon [1ns]			
			Index [2ns]			
			Symbol [3ns]			
		Mediation [2ns]				
		Mentation [3ns]				
Particulars [2ns]						
	MonadicDyads [1ns]					
		MonoidalDyad [1ns]				
		EssentialDyad [2ns]				
		InherentialDyad [3ns]				
	Events [2ns]					
		Spontaneous [1ns]				

(continued)

Table B.1 (continued)

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
		Action [2ns]				
			Exertion [1ns]			
			Perception [2ns]			
			Thought [3ns]			
		Continuous [3ns]				
			TriadicAction [1ns]			
			Activities [2ns]			
			Processes [3ns]			
	Entities [3ns]					
		SingleEntities [1ns]				
			Phenomenal [1ns]			
			States [2ns]			
				Situations		
			Continuants [3ns]			
				Space		
					Points [1ns]	
					Areas [2ns]	
						2D dimensions
					SpaceRegions [3ns]	
						3D dimensions
				Time		
					Instants [1ns]	
					Intervals [2ns]	
					Eternals [3ns]	
		PartOfEntities [2ns]				
			Members [1ns]			
			Parts [2ns]			
			FunctionalComponents [3ns]			
		ComplexEntities [3ns]				
			CollectiveStuff [1ns]			
			MixedStuff [2ns]			
			CompoundEntities [3ns]			
Generals [3ns] (== SuperTypes)						
	Constituents [1ns]					
		NaturalPhenomena [1ns]				
		SpaceTypes [2ns]				
			Shapes [1ns]			
			Places [2ns]			
				LocationPlace		
				AreaRegion		
			Forms [3ns]			
		TimeTypes [3ns]				
			Times [1ns]			

(continued)

Table B.1 (continued)

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
			EventTypes [2ns]			
			ActivityTypes [3ns]			
	Predications [2ns]					
		AttributeTypes [1ns]				
			IntrinsicAttributes [1ns]			
			AdjunctualAttributes [2ns]			
			ContextualAttributes [3ns]			
		RelationTypes [2ns]				
			DirectRelations [1ns]			
			CopulativeRelations [2ns]			
				ActionTypes		
			MediativeRelations [3ns]			
				SituationTypes		
		RepresentationTypes [3ns]				
			Denotatives [1ns]			
			Indexes [2ns]			
			Associatives [3ns]			
	Manifestations [3ns]					
		NaturalMatter [1ns]				
			AtomsElements [1ns]			
			NaturalSubstances [2ns]			
			Chemistry [3ns]			
		OrganicMatter [2ns]				
			OrganicChemistry [1ns]			
				BiologicalProcesses		
			LivingThings [2ns]			
				Prokaryotes [1ns]		
				Eukaryotes [2ns]		
					ProtistsFungus [1ns]	
					Plants [2ns]	
					Animals [3ns]	
				Diseases [3ns]		
			Agents [3ns]			
				Persons [1ns]		
				Organizations [2ns]		
				Geopolitical [3ns]		
		Symbolic [3ns]				
			Information [1ns]			
				AVInfo [1ns]		
					VisualInfo	
					AudioInfo	
				WrittenInfo [2ns]		
				StructuredInfo [3ns]		

(continued)

Table B.1 (continued)

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
			Artifacts [2ns]			
				FoodDrink		
				Drugs		
				Products		
				Facilities		
			Systems [3ns]			
				ConceptualSystems [1ns]		
					Concepts [1ns]	
					TopicsCategories [2ns]	
					LearningProcesses [3ns]	
				SocialSystems [2ns]		
					FinanceEconomy	
					Society	
				Methodeutic [3ns]		
					InquiryMethods [1ns]	
					KnowledgeDomains [2ns]	
					EmergentKnowledge [3ns]	

Note that Table 10.2 in the main book provides an expansion on the typologies found under the Generals branch in the table above.

Capabilities and Uses

Online demos, various search and discovery facilities, use cases, and documentation on the KBpedia Web site provide further details about KBpedia. The primary purpose of KBpedia is to serve as a starting template for creating local domain knowledge graphs. However, as is, KBpedia also has the following capabilities:

- A consistent, coherent combination of six (6) large and leading public knowledge bases into the computable KBpedia knowledge structure
- Mappings to a further 20 'extended' knowledge bases
- A structured organization of the contributing knowledge sources that enables separate treatment of concepts, entities, events, attributes, and relations and their associated types
- Powerful and flexible manipulation and filtering capabilities
- Robust and configurable search and retrieval functions
- Pre-built taggers, classifiers, and mappers
- Ingest and export of multiple data formats
- All functions available via microservice-like APIs and Web services
- Use of open and accepted languages and standards

- A modular and expandable architecture
- A completely Web-based system, which we may deploy locally or in the cloud
- Integration and incorporation of all data assets—unstructured text, semi-structured and markup data, and structured datasets and databases
- A reference structure for interrelating and integrating your domain content
- Inherent multi-linguality, supported by the 200+ languages of the source knowledge bases
- Precise semantic representation for all items, enabling better selections and matches
- The ability to make selections via inference and other logical operations
- The potential to recognize and train up to 47,000 fine-grained entity types
- A knowledge graph suitable for network analytics such as influence, centrality, shortest paths, assortative mixing, and betweenness
- Faster, cheaper creation of positive and negative machine learning training sets
- Faster, cheaper configuration and testing of machine learners.

You may download the open-source KBpedia and other supporting materials and documentation from the project's GitHub site.

References

1. M.K. Bergman, Cognonto is on the hunt for big AI Game, in *AI3:::Adaptive Information* (2016)
2. E. Gabrilovich, S. Markovitch, Computing semantic relatedness using wikipedia-based explicit semantic analysis, in *Proceedings of the 20th International Joint Conference on Artificial Intelligence* (*IJCAI*) (Hyderabad, India, 2007)
3. M. Hassan, *Automatic Document Topic Identification Using Hierarchical Ontology Extracted from Human Background Knowledge* (2013)
4. F. Wu, D.S. Weld, Automatically refining the wikipedia infobox ontology, in *Proceedings of the 17th International Conference on World Wide Web* (ACM, 2008), pp. 635–644
5. A. Kittur, E.H. Chi, B. Suh, What's in wikipedia? Mapping topics and conflict using socially annotated category structure, in *Proceedings of the 27th Annual CHI Conference on Human Factors in Computing Systems (CHI'2009)* (New York, USA, 2009), pp. 1509–1512
6. H. Paulheim, C. Bizer, Type inference on noisy RDF data, in *International Semantic Web Conference* (Springer, Heidelberg, 2013), pp. 510–525
7. D. Vrandečić, M. Krötzsch, Wikidata: A Free Collaborative Knowledgebase. Commun. ACM **57**, 78–85 (2014)

Appendix C: KBpedia Feature Possibilities

The two most labor-intensive steps in machine learning for natural language are (1) feature engineering and (2) labeling of training sets. Supervised machine learning uses an input-output pair, mapping an input, which is a feature, to an output, which is the label. The machine learning consists of inferring ('learning') a function that maps between these inputs and outputs with adequate predictive power. We can apply this learned function to previously unseen inputs to predict the output label. The technique is particularly suited to problems of regression or of classification. Yet, despite the integral role of *features* in the machine learning process, we often overlook their importance compared to labels and algorithms.

Before we can understand how best to leverage features in our *knowledge-based artificial intelligence* (KBAI) efforts, we need first to define and name the feature space. Separately, we also need to study what exists on how to select, construct, extract, or engineer these features. Armed with this background, we can now assemble an inventory of what features might contribute to natural language or knowledge base learning.

We have followed these steps to produce a listing of possible KBpedia features.[1] We have organized that inventory a bit to point out the structural and conceptual relationships among these features, which enables us to provide a lightweight taxonomy for the space. Since others have not named or exposed many of these features before, we conclude this appendix with some discussion about what next-generation learners may gain by working against this structure. Of course, since much of this thinking is incipient, forks and dead ends may unfold, but there also will likely be unforeseen expansions and opportunities as well. A systematic view of machine learning and its knowledge and human language features—coupled with large-scale knowledge bases such as Wikipedia and Wikidata—can lead to faster and cheaper learners across a comprehensive range of NLP tasks.

[1] Features apply to any form of machine learning, including for things like image, speech, and pattern recognition. However, this chapter is limited to the context of natural language, unstructured data, and knowledge bases.

M. K. Bergman, *A Knowledge Representation Practionary*,
https://doi.org/10.1007/978-3-319-98092-8

What Is a Feature?

A "feature is an individual measurable property of a phenomenon being observed" [1]. It is an input to a machine learner, an explanatory variable, sometimes in the form of a function. Some equate features with attributes, but this is not strictly accurate, since a feature may be a combination of other features, or a statistical calculation, or an abstraction of other inputs (some would say it could be about anything!). In any case, we must express a feature as a numeric value (including Boolean as 0 and 1) upon which the machine learner can calculate its predictions. Machine learner predictions of the output can only be based on these numeric features, though they can be subject to rules and weights depending on the type of learner.

Pedro Domingos emphasizes the importance of features and the fact they may be extracted or constructed from other inputs [2]:

> At the end of the day, some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used… Often, the raw data is not in a form that is amenable to learning, but you can construct features from it that are. This is typically where most of the effort in a machine learning project goes. It is often also one of the most interesting parts, where intuition, creativity and 'black art' are as important as the technical stuff.

Many experienced ML researchers make a similar reference to the art or black art of features. In broad strokes in the context of natural language, a feature may be a surface form, like terms or syntax or structure (such as hierarchy or connections); derived (such as statistical, frequency, weighted, or based on the ML model used); semantic (in terms of meanings or relations); or latent, either as something hidden or abstracted from feature layers below it. Unsupervised learning or deep learning features arise from the latent form.

For a given NLP problem domain, features can number into the millions or more. Concept classification, for example, could use features corresponding to all of the unique words or phrases in that domain. Relations between concepts could also be as numerous. We calculate some form of vector relationship over, say, all of the terms in the space so that we may assign a numerical value to 'high-dimensional' features.[2] Because learners may learn about multiple feature types, the potential combinations for the ML learner can be astronomical. This combinatorial problem has been known for decades and has been termed the curse of dimensionality for more than 50 years [3].

Of course, just because a feature exists says nothing about whether it is useful for ML predictions. Features may thus be one of the four kinds: (1) strongly relevant, (2) weakly relevant, (3) irrelevant, or (4) redundant [10]. We should favor strongly relevant features; we may sometimes combine weakly relevant to improve the overall relevancy. We should remove all irrelevant or redundant features from consideration. Often, the fewer the features, the better, so long as the features used are strongly relevant and orthogonal (that is, they capture different aspects of the prediction space) to one another.

[2] For example, in the term or phrase space, the vectors might be constructed from counts, frequencies, cosine relationships between representative documents, distance functions between terms, etc.

A (Partial) Inventory of Natural Language and KB Features

To make this discussion tangible, we have assembled a taxonomy of feature types in the context of natural language and knowledge bases. I drew this inventory from the limited literature on feature engineering and selection in the context of KBAI from the perspectives of ML learning in general [4–6], ML learning ontologies [7, 8], and knowledge bases [9–13]. My listing is only partial, but does provide an inventory of more than 200 feature types applicable to natural language.

I have organized this inventory into eight main areas in Table C.1, shown in *italics*, which tend to cluster into these four groupings:

- ***Surface features****—These are features that one can see within the source documents and knowledge bases. They include****lexical****items for the terms and phrases in the domain corpus and knowledge base;****syntactical****items that show the word order or syntax of the domain;****structural****items that either split the documents and corpus into parts or reflect connections and organizations of the items, such as hierarchies and graphs; or****natural language****items that reflect how we express the content in the surface forms of various human languages.*
- ***Derived features****—These are surface features that we transform or derive in some manner, such as the direct****statistical****items or the****model-based****ones reflecting the characteristics of the machine learners used.*
- ***Semantic features****—These are summarized in the****semantics****area, and reflect what the various items mean or how they are conceptually related to one another.*
- ***Latent features****—These features are not observable from the source content. Instead, these are statistically derived abstractions of the features above that are one- to N-levels removed from the initial source features. These****latent****items may either be individual features or entire layers of abstraction removed from the surface layer. These features result from applying unsupervised or deep learning machine learners.*

We may nucleate features and training sets based on the syntax, morphology, semantics (meaning of the data), or relationships (connections) of the source data in the knowledge base. Continuous testing and application of machine learning to the system itself create virtuous feedback where the accuracy of the overall system is constantly and incrementally improved.

The compiled taxonomy listing of features in Table C.1 exceeds any prior listings. In fact, most of the feature types we show have yet to participate in NLP machine learning tasks. We organize our taxonomy according to the same eight main areas, shown under the shaded entries, noted above:

Fully 50% of the features listed in the inventory in Table C.1 above arise from unique KB aspects, especially in the areas of ***semantics*** and ***structural***, including graph relationships. Many, if not most, of these new feature possibilities may prove redundant or only somewhat relevant or perhaps not at all. Not all features may ever prove useful. Some, such as *Case*, may be effectively employed for named entity or specialty extractions, applicable to copyrights or unique IDs or data types, but may prove of little use in other areas.

Still, because many of these KB features cover orthogonal aspects of the source knowledge bases, the likelihood of finding new, strongly relevant features is high. Further, except for the ***latent*** and ***model-based*** areas, each of these feature types may be used singly or in combination to create coherent slices for both positive and negative training sets, helping to reduce the effort for labor-intensive labeling as well. By extension, we can use these capabilities to more effectively bootstrap the creation of gold standards, useful when we are testing parameters.

The *Statistical* and *Meta-features* sections of Table C.1 are first derivatives of the base structure. The few listed here are examples of how we may include such measures in the feature pool, and they all are common ones. The point is that we may use derivatives and embeddings from other features in the table as legitimate features in their own right.

Table C.1 A (partial) taxonomy of machine learning features

Lexical				
	Corpus			
	Phrases			
		Averages		
		Counts		
		N-grams		
		Weights		
	Words			
		Averages		
		Counts		
		Cutoffs (top *N*)		
		Dictionaries		
		Named entities		
		Stemming		
		Stoplists		
		Terms		
		Weights		
Syntactical				
	Anaphora			
	Cases			
	Complements (argument)			
	Co-references			
	Decorations			
	Dependency grammar			
		Head (linguistic)		

(continued)

Table C.1 (continued)

	Distances			
	Gender			
	Moods			
	Paragraphs			
	Parts of speech (POS)			
	Patterns			
	Plurality			
	Phrases			
	Sentences			
	Tenses			
	Word order			
Statistical				
	Articles			
		Vectors		
	Information-theoretic			
		Entropy		
		Mutual information		
	Meta-features			
		Correlations		
		Eigenvalues		
		Kurtosis		
		Sample measures		
			Accuracy	
			F-1	
				Precision
				Relevance
		Skewness		
		Vectors		
		Weights		
	Phrases			
		Document frequencies		
		Frequencies (corpus)		
		Ranks		
		Vectors		
	Words			
		Document frequencies		
		Frequencies (corpus)		
		Ranks		
		String similarity		
		Vectors		
			Cosine measures	
			Feature vectors	

(continued)

Table C.1 (continued)

Structural				
	Documents			
		Node types		
			Depth	
			Leaf	
	Document parts			
		Abstract		
		Authors		
		Body		
		Captions		
		Dates		
		Headers		
		Images		
		Infoboxes		
		Links		
		Lists		
		Metadata		
		Templates		
		Title		
		Topics		
	Captions			
	Disambiguation pages			
	Discussion pages			
		Authors		
		Body		
		Dates		
		Links		
		Topics		
	Formats			
	Graphs (and ontologies)			
		Acyclic		
		Concepts		
			Centrality	
			Relatedness	
		Directed		
		Metrics (counts, averages, min/max)		
			Attributes	
			Axioms	
			Children	
			Classes	
			Depth	
			Individuals	
			Parents	
		Subgraphs		

(continued)

Table C.1 (continued)

	Headers			
		Content		
		Section hierarchy		
	Infoboxes			
		Attributes		
		Missing attributes		
		Missing values		
		Templates		
		Values		
	Language versions			
		Definitions		
		Entities		
		Labels		
		Links		
		Synsets		
	Links			
		Category		
		Incoming		
		Linked data		
		Outgoing		
		See also		
	Lists			
		Ordered		
		Unordered		
	Media			
		Audio		
		Images		
		Video		
	Metadata			
		Authorship		
		Dates		
		Descriptions		
		Formats		
		Provenance		
	Pagination			
	Patterns			
		Dependency patterns		
		Surface patterns		
			Regular expressions	
	Revisions			
		Authorship		
		Dates		
		Structure		

(continued)

Table C.1 (continued)

			Document parts	
				Captions
				Headers
				Infoboxes
				Links
				Lists
				Metadata
				Templates
				Titles
		Versions		
	Source forms			
		Advertisements		
		Blog posts		
		Documents		
			Research articles	
			Technical documents	
		E-mails		
		Microblogs (tweets)		
		News		
		Technical		
		Web pages		
	Templates			
	Titles			
	Trees			
		Breadth measures		
		Counts		
		Depth measures		
	Web pages			
		Advertisements		
		Body		
		Footer		
		Header		
		Images		
		Lists		
		Menus		
		Metadata		
		Tables		
Semantics	(most also subject to *syntactical* and *statistical* features above)			
	Annotations			
		Alternative labels		
		Notes		
		Preferred labels		
	Associations			
		Association rules		

(continued)

Table C.1 (continued)

		Co-occurrences		
		See also		
	Attribute types			
		Attributes		
			Cardinality	
			Descriptive	
			Qualifiers	
			Quantifiers	
				Many
		Values		
			Data types	
			Many	
	Categories			
		Eponymous pages		
	Concepts			
		Definitions		
		Grouped concepts (topics)		
		Hypernyms		
			Hypernym-based feature vectors	
		Hyponyms		
		Meanings		
		Synsets		
			Acronyms	
			Epithets	
			Jargon	
			Misspellings	
			Nicknames	
			Pseudonyms	
			Redirects	
			Synonyms	
	Entity types			
		Entities		
		Events		
		Locations		
	General semantic feature vectors			
	Relation types			
		Binary		
		Identity		
		Logical conjunctions		
			Conjunctive	
			Disjunctive	
		Mereology (part of)		
		Relations		
			Domain	
			Range	

(continued)

Table C.1 (continued)

		Similarity		
	Roles			
	Voice			
		Active/passive		
		Gender		
		Mood		
		Sentiment		
		Style		
		Viewpoint (worldview)		
Natural languages				
	Morphology			
	Nouns			
	Syntax			
	Verbs			
	Word order			
Latent				
	Autoencoders			
		Many; dependent on method		
	Features			
		Many; dependent on method		
	Hidden			
		Many; dependent on method		
	Kernels			
		Many; dependent on method		
Model based				
	Decision tree			
		Tree measures		
	Dimensionality			
	Feature characteristics			
		Data types		
		Max		
		Mean		
		Min		
		Number		
		Outliers		
		Standard deviation		
	Functions			
		Factor graphs		
		Functors		
		Mappings		
	Landmarking			
		Learner accuracy		
	Method measures			
		Error rates		

Though the literature most often points to classification as the primary use of knowledge bases as background knowledge supporting machine learners, in fact, many natural language processing (NLP) tasks may leverage KBs. Here is but a brief listing of application areas for KBAI:

See also Table 4.1. Undoubtedly other applications will emerge as this more systematic KBAI approach to machine learning evolves over the coming years.

Feature Engineering for Practical Limits

This richness of feature types leads to the combinatorial problem of too many features. Feature engineering is the way both to help find the features of strongest relevance and to reduce the feature space dimensionality to speed the ML learning times. Initial feature engineering tasks should be to transform input data, regularize them if need be, and create numeric vectors for new ones. These are preparation tasks to convert the source or target data to forms amenable to machine learning. This staging now enables us to discover the most relevant ('strong') features for the given ML method under investigation.

In a KB context, specific learning tasks as outlined in Table C.2 are often highly patterned. The most effective features for training, say, an entity recognizer, will only involve a limited number of strongly relevant feature types. Moreover, the relevant feature types applicable to a given entity type should mostly apply to other entity types, even though the specific weights and individual features (attributes and other type relations) will differ. This patterned aspect means that once we train a given ML learner for a given entity type, its relevant feature types should be approximately applicable to other related entity types. We can reduce the lengthy process of initial feature selection as training proceeds for similar types. It appears that we may discover combinations of feature types, specific ML learners, and methods to create training sets and gold standards for entire classes of learning tasks.

Table C.2 NLP applications for machine learners using KBs

• Entity recognizers	• Ontology development
• Relation extractors	• Entity dictionaries
• Classifiers	• Data conversion and mapping
• Q & A systems	• Master data management
• Knowledge base mappings	• Specialty extractors

Probably the most time-consuming and demanding aspect of these patterned approaches resides in *feature selection* and *feature extraction*. *Feature selection* is the process of finding a subset of the available feature types that provide the highest predictive value while not overfitting.[3] Researchers typically split feature selection into three main approaches [14, 15, 16]:

[3] Overfitting is where a statistical model, such as a machine learner, describes random error or noise instead of the underlying relationship. It is particularly a problem in high-dimensional spaces, a common outcome of employing too many features.

- Filter—Select the N most promising features based on a ranking from some form of proxy measure, like *mutual information* or the *Pearson correlation coefficient*, which provides a measure of the information gain from using a given feature type.
- Wrapper—Test feature subsets through a greedy search heuristic that either starts with an empty set and adds features (forward selection) keeping the "strongest" ones or starts with a full set and gradually removes the "weakest" ones (backward selection); the wrapper approach may be computationally expensive.
- Embedded—Include feature selection as a part of model construction.

For high-dimensional features, such as terms and term vectors, we may apply stoplists or cutoffs (only considering the top *N* most frequent terms, for example) to reduce dimensionality. Part of the 'art' portion resides in knowing which feature candidates may warrant formal selection or not; this learning can be codified and reused for similar applications. One may also apply some unsupervised learning tests at this point to discover additional 'strong' features.

Feature extraction transforms the data in the high-dimensional space to a space of fewer dimensions. Functions create new features in the form of ***latent*** variables, which are not directly observable. Also, because these are statistically derived values, many input features are reduced to the synthetic measure, which naturally causes a reduction in dimensionality. Advantages of a reduction in dimensionality include:

1. Often a better feature set (resulting in better predictions) [17]
2. Faster computation and smaller storage
3. Reduction in collinearity due to a reduction in weakly interacting inputs
4. Easier graphing and visualization.

On the other hand, the latent features are abstractions, and so not easily understood as the literal. Deep learning generates multiple layers of these latent features as the system learns.

Of course, we may also combine the predictions from multiple ML methods, which then also raises the questions of ensemble scoring. We may also self-learn (that is, meta-learn) more systematic approaches to ML such that the overall learning process can proceed in a more automated way.

Considerations for a Feature Science

In supervised learning, it is clear that more time and attention have been given to the labeling of the data, what the desired output of the model should be. Much less time and attention have been devoted to features, the input side of the equation. The purposeful use of knowledge bases and structuring them is one way we can make progress. Still, progress also requires some answers to some fundamental questions. A scientific approach to the feature space would likely need to consider, among other objectives:

- Full understanding of surface, derived, and latent features
- Relating various use cases and problems to specific machine learners and classes of learners
- Relating specific machine learners to the usefulness of particular features (see also *hyperparameter optimization* and *model selection*)
- Improved methods for feature engineering and construction
- Improved methods for feature selection
- A better understanding of how to select supervised and unsupervised ML.

Some tools and utilities would also help to promote this progress. Some of these capabilities include:

- Feature inventories—how to create and document taxonomies of feature types
- Feature generation—methods for the codification of leading recipes
- Feature transformations—the same for transformations, up to and including vector creation.

Role of a Platform

The object of these efforts is to systematize how knowledge bases, combined with machine learners, can speed the deployment and lower the cost of creating bespoke artificial intelligence applications of natural language for specific domains. KBAI places primary importance on *features*. An abundance of opportunity exists in this area, and an abundance of work required, but little systematization.

The good news is we can build platforms that manage and grow the knowledge bases and knowledge graphs supporting machine learning, as we discussed in *Parts III* and *IV*. We can apply machine learners in a pipeline manner to these KBs, including orchestrating the data flows in generating and testing features, running and testing learners, creating positive and negative training sets, and establishing gold standards. The heart of the platform must be an appropriately structured knowledge base organized according to a coherent knowledge graph; this is the primary purpose of KBpedia.

Still, in the real world, engagements always demand unique scope and unique use cases. We should engineer our platforms to enable ready access, extensions, configurations, and learners. It is vital to structure our source knowledge bases such that slices and modules can be specified, and all surface attributes may be selected and queried. Mapping to the external schema is also essential. Background knowledge from a coherent knowledge base is the most efficient way to fuel this.

References

1. C. Bishop, *Pattern Recognition and Machine Learning* (Springer, Berlin, 2006)
2. P. Domingos, A Few Useful Things to Know About Machine Learning. Commun. ACM **55**, 78–87 (2012)
3. R.E. Bellman, *Dynamic Programming* (Princeton University Press, Rand Corporation, 1957)
4. I. Guyon, A. Elisseeff, An introduction to feature extraction, in *Feature Extraction* (2006), pp. 1–25
5. D. Haussler, *Convolution Kernels on Discrete Structures* (Technical report, Department of Computer Science, University of California at Santa Cruz, Irvine, CA, 1999)
6. M. Reif, F. Shafait, M. Goldstein, T. Breuel, A. Dengel, Automatic Classifier Selection for Non-Experts. Pattern. Anal. Applic. **17**, 83–96 (2014)
7. J. Tang, S. Alelyani, H. Liu, Feature selection for classification: a review, in *Data Classification: Algorithms and Applications* (2014), p. 37
8. M. Hilario, P. Nguyen, H. Do, A. Woznica, A. Kalousis, Ontology-based meta-mining of knowledge discovery workflows, in *Meta-Learning in Computational Intelligence* (Springer Berlin, Heidelberg, 2011), pp. 273–31
9. P. Panov, L. Soldatova, S. Džeroski, Ontology of Core Data Mining Entities. Data Min. Knowl. Disc. **28**, 1222–1265 (2014)
10. I. Anastacio, B. Martins, P. Calado, Supervised learning for linking named entities to knowledge base entries, in *Proceedings of the Text Analysis Conference (TAC2011)* (2011)
11. W. Cheng, G. Kasneci, T. Graepel, D. Stern, R. Herbrich, Automated feature generation from structured knowledge, in *Proceedings of the 20th ACM International Conference on Information and Knowledge Management* (ACM, 2011), pp. 1395–1404
12. L. Huang, D. Milne, E. Frank, I.H. Witten, Learning a Concept-Based Document Similarity Measure. J. Assoc. Inf. Sci. Technol. **63**, 1593–1608 (2012)
13. O. Medelyan, D. Milne, C. Legg, I.H. Witten, Mining Meaning from Wikipedia. Int. J. Hum. Comput. Stud. **67**, 716–754 (2009)
14. H. Shen, M. Chen, R. Bunescu, R. Mihalcea, Wikipedia Taxonomic Relation Extraction using Wikipedia Distant Supervision. Ann. Arbor. **1001**, 48109 (2014)
15. G.H. John, R. Kohavi, K. Pfleger, Irrelevant features and the subset selection problem, in *Machine Learning: Proceedings of the Eleventh International Conference* (1994), pp. 121–129
16. Z. Žabokrtský, *Feature Engineering in Machine Learning* (2015)
17. G.E. Hinton, Deep Belief Networks. *Scholarpedia***4**, 5947 (2009)

Glossary

ABox An ABox (for assertions, the basis for A in ABox) is an 'assertion component'; that is, a fact associated with a terminological vocabulary within a knowledge base. ABox are *TBox*-compliant statements about instances belonging to the concept of an ontology. Instances and instance records reside within the Abox.

Abductive reasoning Abduction (or abductive reasoning) is a mode of symbolic inference that involves the screening and selection from a domain D of the possible explanation paths to an outcome O possibly involving any element E of D, with the selection of candidate paths for inductive testing based on plausibility, economy, and potential impact; abduction does not produce probable results, only winnowed candidates.

Access control Access control is the protection of resources against unauthorized access, a process by which use of resources is regulated according to a security policy and is permitted by only authorized system entities according to that policy; see further RFC 2828.

Actions Actions are reactions to perceptions or stimuli, are energetic, or are thought, as understood to be broadly construed; actions reside in *Secondness*.

Activities Activities are sustained actions over durations of time; activities may be organized into natural classes.

Accuracy It is a statistical measure of how well a binary classification test correctly identifies or excludes a condition. It is calculated as the sum of *true positives* and *true negatives* divided by the total population.

Adaptive ontology An adaptive ontology is a conventional knowledge representational ontology that has added to it a number of specific best practices, including modeling the *ABox* and *TBox* constructs separately; information that relates specific types to different and appropriate display templates or visualization components; use of *preferred labels* for user interfaces, as well as alternative labels and hidden labels; defined concepts; and a design that adheres to the *open-world assumption*

M. K. Bergman, *A Knowledge Representation Practionary*,
https://doi.org/10.1007/978-3-319-98092-8

Administrative ontology Administrative ontologies govern internal application use and user interface interactions.

Annotation Annotations are indexes and the *metadata* of the *KB*; these cannot be inferenced over and do not participate in *reasoning* or *coherency* testing. But, they can be searched, and language *features* can be processed in other ways. Annotations may be grouped for convenience, but may not be typed.

API An application programming interface (API) defines how communication may take place between applications. Implementing APIs that are independent of a particular operating environment (as are the W3C DOM Level 2 specifications) may reduce implementation costs for multi-platform user agents and promote the development of multi-platform assistive technologies.

Architectonic Architectonic is a governing philosophy expressed iconically as a design, schema, or structure.

Architecture Architecture, as limited to the use herein, is the structure of a knowledge base (or bases) written in a knowledge representation formalism, and embedded in a general knowledge management platform. The architecture combines knowledge artifacts with software program(s) or computing system(s), the relationships among them, and the conditions on their use.

Artificial intelligence AI is the use of computers to do or assist complex human tasks or *reasoning*. There are many broad subfields from pattern recognition to robotics and complex planning and optimizations.

Aspects Aspects are aggregations of an *entity type* that are grouped according to *features* or views different from the type itself. As examples, the type of 'music composer' may have an aspect of being from the nineteenth century, or 'authors' may have the aspects of being Russian or writing novellas. The organization of aspects closely parallels that for *SuperTypes*.

Assertion Assertion is a *statement*, wherein a fact or logical expression with consequences is made.

Attributes *Attributes* are the *intensional* characteristics of an *object*, *event*, *entity*, *type* (when viewed as an *instance*), or *concept*. The relationship is between the individual particular and its attributes and characteristics, in the form of A:A. Attributes may be intrinsic characteristics or essences of single particulars, adjunctual or accidental happenings to the particular, or contextual in time or space or situations. Collectively known as depth, comprehension, significance, meaning, or connotation.

Attribute type Attribute types are an aggregation (or class) of multiple *attributes* that have similar characteristics among themselves (for example, colors or ranks or metrics). As with other types, attribute types do not have attributes.

Axiom An axiom is a premise or starting point of *reasoning*. In an *ontology*, each statement (*assertion*) is an axiom.

Base concept Base concepts are all of the *classes* in the overall KBpedia, comprised of the KBpedia Knowledge Ontology and all of its official *typologies*.

Belief Belief is a state of evidence sufficient to enable action.

Binding Binding is the creation of a simple reference to something that is larger and more complicated and used frequently. The simple reference can be used instead of having to repeat the larger thing.

Blank node Also called a bnode, a blank node in *RDF* is a *resource* for which a *URI* or literal is not given. A blank node indicates the existence of a thing, implied by the structure of the *knowledge graph*, but which was never explicitly identified by giving it a URI. Blank nodes have no meaning outside of their current graph and therefore cannot be mapped to other resources or graphs.

Cardinality Cardinality is where the number of members in a class or type is set or limited, such as `hasBiologicalParent` being set to two.

Class Class is a collection of one or more *instances* or *classes* that share the same potential *attributes* or *relations*; *concepts* and *entity types* are both classes.

Closed-world assumption CWA is the premise that what is not currently known to be true is false. CWA is the most common logic applied to relational database systems, useful for transaction-type systems. In knowledge management, CWA is used in at least two situations: (1) when the *knowledge base* is known to be complete, and (2) when the knowledge base is known to be incomplete, but a 'best' definite answer must be derived from incomplete information. See contrast to the *open-world assumption*.

Coherence Coherence is the state of being coherent for logic systems, where the knowledge base (domain) is consistent and has a high degree of conjunction for nondeductive assertions.

Collection See *class*.

Complete Complete is an evaluative criterion for logic systems where all *statements* which are true in all models are provable.

Concept See *class*.

Cyc Cyc is a common-sense *knowledge base* that has been under development for over 20 years backed by 1000 person-years of effort. The smaller OpenCyc version is available in OWL as open source; a ResearchCyc version of the entire system is available to researchers. The Cyc platform contains its own logic language, CycL, and has many built-in functions in areas such as *natural language processing*, search, inferencing, and the like. *UMBEL* is based on a subset of Cyc.

Cycle A cycle is where, in a graph, a path from a given *node* may reach itself (such as A → B → C → A). In a subsumption hierarchy, this is an error.

Data integration Data integration is the bringing together of data from heterogeneous and often physically distributed data sources into a single, coherent view.

Dataset Dataset is a combination of one or more *records*, transmitted as a single unit (though it may be split into parts due to size).

Data types Data types are predefined ways that attribute values may be expressed, including various literals and strings (by language), *URIs*, Booleans, numbers, date-times, etc.

DBpedia A project that extracts structured content from *Wikipedia*, and then makes that data available as *linked data*. There are millions of entities characterized by

DBpedia in this way. As such, DBpedia is one of the largest—and most central—hubs for *linked data* on the Web.

Deductive reasoning Deductive reasoning extends from premises known to be true and clear to infer new facts.

Description logics Description logics and their semantics traditionally split *concepts* and their *relationship*s from the different treatment of *instances* and their *attributes* and roles expressed as *fact assertions*. The concept split is known as the *TBox* and represents the schema or *taxonomy* of the *domain* at hand. The TBox is the structural and *intensional* component of conceptual relationships. The second split of instances is known as the *ABox* and describes the attributes of instances (and individuals), the roles between instances, and other assertions about instances regarding their class membership with the TBox concepts.

Disjoint Disjoint is where membership in one *class* specifically excludes membership in another; this is a useful *property* in that it allows large, well-designed *knowledge bases* to be 'sliced and diced' for more effective processing or analysis.

Distant supervision A method to use *knowledge bases* to label *entities* automatically in text through *machine learning*, which is then used to extract features and train a machine learning classifier. The knowledge bases provide coherent positive training examples and avoid the high cost and effort of manual labeling.

Documents Documents (or articles or *records*) may be in the form of articles or data records. Whatever the form, extractions are needed to convert source information into the *triples* useful to the *knowledge base*.

Domain Domain is the bounded scope of real-world considerations that may contribute to the *knowledge representation* or information queries at hand. Scoping the domain is one of the first activities undertaken in a new KR project.

Domain (property) Domain, as applied to a property, is a statement that declares the classes or types from which the subject of the assertion must be drawn.

Domain ontology Domain (or content) *ontologies* embody more of the traditional ontology functions such as information interoperability, inferencing, reasoning, and conceptual and knowledge capture of the applicable domain.

Edges Edges, in a graphical representation of a *knowledge graph*, are the connections or the *relations* between *subjects* and *objects*; edges are the linking *properties* in a 'triple' *statement*.

Entailment Entailment is a consequence arising from a statement deemed to be true based on some underlying logic. The logical consequence is said to be *necessary* and *formal*; necessary, because of the rules of the logic (the conclusion is the consequent of the premises); and formal because the logical form of the statements and arguments hold true without regard to the specific *instance* or content.

Entity Entities are the basic, real things in our domain of interest; they are nameable things or ideas that have an identity, are defined in some manner, can be referenced, and may be related to *types*; entities are most often the bulk of an overall *knowledge base*. An entity is an *individual object* or *instance* of a *class*, a *Secondness*; when affixed with a proper name or label it is also known as a *named entity* (thus, named entities are a subset of all entities). Entities are described and characterized by *attributes*. Entities are connected or related to one another through *external relations* and are referred to, signified, or indexed by *representations*. An entity may be independent or separate or can be part of something else. Entities cannot be *topics* or *types*.

Entity recognition The use of *natural language processing* to identify specific *entities* in text. Often used in conjunction with named entities, where it is abbreviated NER.

Entity type *Entity types* are the aggregations or collections or *classes* of similar entities, which also share some essence; entity types have the *attributes* (but not the same values) of instances of the type.

Essence The *attribute* or set of attributes that make an entity what it fundamentally is; it is a unique or distinguishing attribute that helps define a *type*.

Event Events are nameable sequences of time, are described in some manner, can be referenced, and may be related to other time sequences or types. Events are like entities, except that they have a discrete time beginning and end. Events are a Secondness and may be typed.

Extensional The extension of a *class*, concept, idea, or sign consists of the things to which it applies, in contrast with its *intension*. For example, the extension of the word 'dog' is the set of all (past, present, and future) dogs in the world. The extension is most akin to the *attributes* or characteristics of the *instances* in a set defining its class membership.

External linkages External linkages (or *mappings*) are any of the relational properties that may be used to map external datasets and schema to KBpedia. In its base form, which can be expanded, KBpedia has mappings to more than 20 external sources.

External relations External relations are assertions (relationships) between an *object*, *event*, *entity*, *type*, or *concept* and another particular or general. An external relationship has the form of *A:B*. External relations may be simple ones of a direct relationship between two different instances; may be copulative by combining objects or asserting membership, quantity, action, or circumstance; or mediative to provide meaning, context, relevance, generalizations, or other explanations of the subject with respect to the external world. External relations are extensional.

Fact A basic *statement* or *assertion* within a *knowledge graph* or *knowledge base*.

Fallibility Fallibility is the doctrine that truth is a limit function, unknowable in the absolute, and provides the logical basis for questioning premises.

False negative An error where a test result indicates that a condition failed, while it actually was successful. That is, the test result indicates a negative when the correct result should have been positive. Also known as a false-negative error or *Type II error* in statistics. It is abbreviated FN.

False positive An error where a test result indicates that a condition was met or achieved, while it actually should have failed. That is, the test result indicates a positive when the correct result should have been negative. Also known as a false-positive error or *Type I error* in statistics. It is abbreviated FP.

Feature A feature is a measurable property of the system being analyzed, equivalent to what is known as an explanatory variable in statistics.

Feature engineering Feature engineering is a process of creating, generating, and selecting the *features* to be used in *machine learning*, based on an understanding of the underlying data and choosing features based on their likely impact on learning results and effectiveness.

Firstness Firstness is possibility, the essences of what may be, the unexpected chance occurrence.

Folksonomy A folksonomy is a user-generated set of open-ended labels called *tags* organized in some manner and used to categorize and retrieve Web content such as Web pages, photographs, and Web links.

Function Function is any algebraic or logical expression allowable by the semantics and primitives used in the KR language where an input is related to an output.

Generals Generals are the mediating, continuous, vague, and indeterminate aggregations of instances into concepts, *classes*, *types*, collections, or sets. Generals are in *Thirdness*. Generals may often be considered real, and their understanding and identification may be shared through *representations*.

GeoNames GeoNames integrates geographical data such as names of places in various languages, elevation, population, and others from various sources.

Gold standard A gold standard is a reference, benchmark test set where the results are already scored and known, with a minimum (if not zero) amount of *false positives* or *false negatives*; good gold standards also include *true-negative* results.

Identifier Identifier is a reference pointer, a sign pointing to an *object*, but not the *object* itself. Identifiers should not be confused with the naming or defining label for the *object*; in practice, it is often a unique string assigned to the object. In *RDF* and KBpedia this identifier is a *URI*

Individual Individual in RDF and OWL (indeed, commonly in description logics) is synonymous with an *instance* or *entity*; we try not to use this term because of general terminological confusion; see *instance*.

Inductive reasoning A method of reasoning where lines of possible evidence are weighed to determine probable outcomes.

Inference Inference is the act or process of deriving logical conclusions from premises known or assumed to be true. The logic within and between *statements* in an *ontology* is the basis for inferring new conclusions from it, using software applications known as inference engines or *reasoners*.

Instance Instances are individual *entities* or *events*, the ground-level components of a *knowledge base*. Instances may include concrete *objects* such as people, animals, tables, automobiles, molecules, and planets, as well as abstract instances such as numbers and words; instances are in Secondness. An instance is also known as an *individual*, with *member* and *entity* also used somewhat interchangeably.

Instance record An *instance* with one or more *attributes* also provided.

Intensional The intension of a class is what is intended as a definition of what characteristics or properties its members should have. Intension is most akin to the *attributes* or characteristics of the *instances* in a set defining its class membership. It is therefore like the key-attribute pair aspects of an instance (or *ABox*) in an *ontology*.

Inverse Inverse is when a *property*, say, `hasParent`, can be defined as the inverse property of `hasChild`.

Key-value pair Also known as a *name-value pair* or *attribute-value pair*, a key-value pair is a fundamental, open-ended data representation. All or part of the data model may be expressed as a collection of tuples `<attribute name, value>` where each element is a key-value pair. The key is the defined *attribute*, and the value may be a reference to another object or a literal string or value. In *RDF triple* terms, the subject is implied in a key-value pair by nature of the *instance record* at hand.

Kind Used synonymously herein with *class*.

Knowledge base A knowledge base (abbreviated KB or kb) is a special kind of database for knowledge management. As used in KBpedia, a KB includes *instances* and *classes* related to each other via triple statements.

Knowledge-based artificial intelligence Knowledge-based artificial intelligence, or KBAI, is the use of large statistical or *knowledge bases* to inform *feature* selection for machine-based learning algorithms used in AI.

Knowledge graph See *ontology*.

Knowledge management Knowledge management, or KM, is the practice of creating, sharing, finding, annotating, connecting, and extending information and knowledge for a given *domain*.

Knowledge representation A field of *artificial intelligence* dedicated to representing information about the world in a form that a computer system can utilize to solve complex tasks.

Knowledge supervision A method of *machine learning* to use *knowledge bases* in a purposeful way to create features, and negative and positive *training sets* in order to train the classifiers or extractors. *Distant supervision* also uses knowledge bases, but not in such a purposeful, directed manner across multiple machine learning problems.

Leaf nodes Leaf nodes are terminal nodes in a tree structure, often representing instances (but not always so).

Linkage A specification that relates an *object* or *attribute* name to its full *URI* (as required in the *RDF* language).

Linked data Linked data is a set of best practices for publishing and deploying *instance* and *class* data using the *RDF* data model, and uses *uniform resource identifiers* (URIs) to name the data objects. The approach exposes the data for access via the HTTP protocol while emphasizing data interconnections, interrelationships, and context useful to both humans and machine agents.

Lists Lists are unordered *members* or *instances*, with or without gaps or duplicates, useful for bulk assignment purposes. Lists generally occur through a direct relation assignment (*e.g.*, `rdf:Bag`). See *Sequences*.

Machine learning The construction of algorithms that can learn from and make predictions on data by building a model from example inputs. A wide variety of techniques and algorithms ranging from *supervised* to *unsupervised* may be employed.

Mapping A considered correlation of *objects* in two different sources to one another, with the relation between the objects defined via a specific *property*. Linkage is a subset of possible mappings.

Member Used synonymously herein with *instance*.

Metadata Metadata are *annotations* and provide information about one or more aspects of the content at hand such as means of creation, purpose, when created or modified, author or provenance, where located, topic or subject matter, standards used, or other descriptive characteristics. In contrast to an *attribute*, which is an individual characteristic intrinsic to an *instance*, metadata is a description about that data.

Metamodeling Metamodeling is the analysis, construction, and development of the frames, rules, constraints, models, and theories applicable and useful for modeling a predefined class of problems.

Microdata Microdata is a proposed specification used to nest semantics within existing content on web pages. Microdata is an attempt to provide a simpler way of annotating HTML elements with machine-readable tags than the similar approaches of using *RDFa* or *microformats*.

Microformats A microformat (sometimes abbreviated μF or uF) is a piece of markup that allows expression of semantics in an HTML (or XHTML) web page. Programs can extract meaning from a web page that is marked up with one or more microformats.

Natural language processing NLP is the process of a computer extracting meaningful information from natural language input and/or producing natural language output. NLP is one method for assigning structured data characterizations to text content. NLP applications include automatic summarization, coreference resolution, machine translation, named entity recognition, question answering, relationship extraction, topic segmentation and recognition, word segmentation, and word sense disambiguation.

Named entity See *entity*.

Named entity recognition See *entity recognition*; also called NER.

Negation Negation is a unary operation that produces a value of *true* when its operand is false and a value of *false* when its operand is true.

Nodes Nodes, in a graphical representation of a *knowledge graph*, are the *subjects* and *objects* in a '*triple*' *statement*; they are connected to one another via *relations* (or *edges* in a graphical representation).

OBIE Information extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents. Ontology-based information extraction (OBIE) is the use of an ontology to inform a 'tagger' or information extraction program when doing natural language processing. Input ontologies thus become the basis for generating metadata tags when tagging text or documents.

Object An object is an *entity*, *event*, *class*, *concept*, or *property* that can be referred to via an *identifier* of some sort; in KBpedia, every object has a URI *identifier*.

Ontology An ontology is a data model that represents a set of *concepts* or *instances* within a *domain* and the *relationships* between those concepts. Loosely defined, ontologies on the Web can have a broad range of formalism, or expressiveness or *reasoning* power.

Ontology-driven application Ontology-driven applications (or ODapps) are modular, generic software applications designed to operate in accordance with the specifications contained in one or more *ontologies*. The relationships and structure of the information driving these applications are based on the standard functions and roles of ontologies (namely as *domain ontologies*), as supplemented by UI and instruction sets and validations and rules.

Open Semantic Framework The open semantic framework, or OSF, is a combination of a layered architecture and an open-source, modular software stack. The stack combines many leading third-party software packages with open-source *semantic technology* developments from structured dynamics.

Open-world assumption OWA is a formal logic assumption that the truth value of a *statement* is independent of whether or not it is known by any single observer or agent to be true. OWA limits the kinds of *inference* and deductions to those that are known to be true. OWA is useful when we represent knowledge within a system as we discover it, and where we cannot guarantee that we have discovered or will discover complete information, typical of knowledge. See contrast to the *closed-world assumption*.

OWL The Web Ontology Language (OWL) is designed for defining and instantiating formal Web *ontologies*. An OWL ontology may include descriptions of *classes*, along with their related *properties* and *instances*. There are also a variety of OWL dialects.

Particulars Particulars are all *entities* and *events*; they are in *Secondness*.

Pragmatic maxim Pragmatic maxim is the understanding of a topic or an object by an apprehension of all of the practical consequences potentially arising from it.

Pragmatism Pragmatism, what Peirce came to term pragmaticism because of what he felt was a misappropriation of his idea, is the consideration and weighing of available alternatives or explanations in order to pick the most likely ones with a return.

Precision The fraction of retrieved documents that are relevant to the query. It is measured as *true positives* divided by all measured positives (true and false). High precision indicates a high percentage of true positives in relation to all positive results.

Predicate See *Property*.

Preferred label Preferred label (or *prefLabels* or *title*) is the readable string (name) for each *object* in KBpedia. The labels are provided as a readable convenience; the actual definition of the object comes from the totality of its description, prefLabel, altLabels, and connections (placement) within the *knowledge graph*. Labels of all kinds are *representations* and reside in *Thirdness*.

Prescission Prescission, or its verbs *prescind* or *prescinded from*, is the process of comparing two items and seeing if either may exist independent of the other.

If so, we say that the independent one is prescinded from the dependent one; it is one way to determine a subsumption relationship.

Property Property is an official term in *RDF* and *OWL* that includes what we term *attributes*, *external relations*, and *representations*; *w*e try to use the term sparingly, generally when only referencing those items in relation to RDF or OWL.

Punning In computer science, punning refers to a programming technique that subverts or circumvents the type system of a programming language, by allowing a value of a certain type to be manipulated as a value of a different type. When used for *ontologies*, it means to treat a thing as both a *class* and an *instance*, with the use depending on context.

Query Query is a request for information from an agent using a suitable knowledge representation.

Range (property) Range (property) declares the *classes* or data types from which the *object* data or *types* of an *assertion* must be drawn.

RDF Resource Description Framework (RDF) is a data model with a syntax that allows *statements* about *resources* in the form of *subject-predicate-object* expressions, called *triples* in RDF terminology. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object.

RDFa RDFa uses attributes from meta and link elements and generalizes them so that they are usable on all elements allowing annotation markup with semantics. RDFa 1.1 is a W3C Recommendation that removes prior dependence on the XML namespace and expands HTML5 and SVG support, among other changes.

RDF Schema RDFS or RDF Schema is an extensible *knowledge representation* language, providing basic elements for the description of *ontologies*, otherwise called RDF vocabularies, intended to structure RDF resources.

Reasoner A semantic reasoner, reasoning engine, rules engine, or simply a reasoner, is a piece of software able to infer logical consequences from a set of asserted facts or *axioms*. The notion of a semantic reasoner generalizes that of an inference engine, by providing a richer set of mechanisms.

Reasoning Reasoning is one of many logical tests using *inference* rules as commonly specified by means of an ontology language, and often a description language. Many reasoners use first-order predicate logic to perform reasoning; inference commonly proceeds by forward chaining or backward chaining.

Recall The fraction of the documents that are relevant to the query that is successfully retrieved. It is measured as *true positives* divided by all potential positives that could be returned from the corpus. High recall indicates a high yield in obtaining relevant results.

Record As used herein, a shorthand reference to an *instance record*.

Reference concept Reference concepts (or *RefConcepts* or *RCs*), the *base concepts* in KBpedia, are any of the noun *objects* within KBpedia and abbreviated as RC. An RC may be either an *entity*, *entity type*, *attribute*, *attribute type*, *relation*, *relation type*, *topic*, or abstract *concept*. RCs are a distinct subset of the more broadly understood ‘concept’ such as used in the *SKOS RDFS* controlled vocabulary or formal concept analysis or the very general or abstract concepts common to some ontologies. The KBpedia *knowledge graph* is a coherently organized structure of the nearly 60,000 RCs in KBpedia. All RCs are OWL *classes*.

Referent The *object* referred to by an *identifier*.

Reflexivity Reflexivity is when every element of X is related to itself, every *class* is its own subclass, such as every person is a person.

Reinforcement learning Reinforcement learning (RL) is a method of *machine learning* wherein actions are evaluated, most often as a Markov decision process, in accordance with stated performance objectives via a reward function to help converge to those desired goals.

Relation Relations are the way we describe connections between two or more things; *attributes*, *external relations*, and *representations* are all *instances* of the relations *class*.

Relation type An aggregation (or *class*) of multiple *relations* that have similar characteristics among themselves. As with other *types*, shared characteristics are subsumed over some *essence*(s) that give the *type* its unique character.

Representations Representations are signs, and the means by which we point to, draw or direct attention to, or designate, denote, or describe a particular *object*, *entity*, *event*, *type*, or *general*. A representational relationship has the form of *re:A*. Representations can be designative of the subject, that is, be icons or symbols (including labels, definitions, and descriptions); indexes that more or less help situate or provide a traceable reference to the subject; or associations, resemblances, and likelihoods in relation to the subject, more often of an indeterminate character.

Root Root is the name for the top-level node in a *taxonomy*, *knowledge graph*, *ontology*, or *typology*.

RSS RSS (an acronym for Really Simple Syndication) is a family of web feed formats used to publish frequently updated digital content, such as blogs, news feeds, or podcasts.

Satisfies Satisfies means that all *statements* have an interpretation that can be shown to be true.

schema.org schema.org is an initiative launched by major search engines to create and support a common set of schema for structured data markup on web pages. schema.org provided a starter set of schema and extension mechanisms for adding to them. schema.org supports markup in *microdata*, *microformat*, and RDFa formats.

Secondness Secondness is one of the three universal categories and refers to all actual *instances*, specifically including all individual *events* and *entities*.

Semantic enterprise An organization that uses *semantic technologies* and the languages and standards of the *semantic Web*, including *RDF*, *RDFS*, *OWL*, *SPARQL*, and others to integrate existing information assets, using the best practices of *linked data* and the *open-world assumption*, and targeting knowledge management applications.

Semantic technology Semantic technologies combine software and semantic specifications to encode meanings separate from data and content files and separate from application code. This approach enables machines as well as people to understand, share, and reason with data and specifications separately. Semantic technologies provide an abstraction layer above existing IT technologies that enables bridging and interconnection of data, content, and processes.

Semantic Web The Semantic Web is a collaborative movement led by the World Wide Web Consortium (W3C) that promotes common formats for data on the World Wide Web. By encouraging the inclusion of semantic content in web pages, the Semantic Web aims at converting the current web of unstructured documents into a 'web of data.' It builds on the W3C's Resource Description Framework (*RDF*).

Semset Semsets (or *synsets* or *alternative labels* or *altLabels*) are collections of alternate labels and terms to describe a *concept* or *entity* or *event*. These semset alternatives include true synonyms, but may also be more expansive and include abbreviations, acronyms, aliases, argot, buzzwords, cognomens, derogatives, diminutives, epithets, hypocorisms, idioms, jargon, lingo, metonyms, misspellings, nicknames, nonstandard terms (see Twitter), pejoratives, pen names, pseudonyms, redirects, slang, sobriquets, or stage names, in short, any term or phrase that can be a reference to a given *instance* or *class*.

Sequences Sequences are ordered *members* or *instances*, with or without gaps or duplicates, useful for bulk assignment purposes. Sequences generally occur through a direct relation assignment (*e.g.*, `rdf:Seq`). See *Lists*.

SKOS SKOS or Simple Knowledge Organization System is a family of formal languages designed for representation of thesauri, classification schemes, *taxonomies*, subject-heading systems, or any other type of structured controlled vocabulary; it is built upon *RDF* and *RDFS*.

Sound An evaluative criterion in logic systems where all provable statements are true in all models.

SPARQL SPARQL (pronounced 'sparkle') is an *RDF* query language; its name is a recursive acronym that stands for SPARQL Protocol and RDF Query Language.

Statement A statement is the standard and most basic expression in *RDF* and *OWL*. A statement is comprised of a '*triple*' (*subject-property-object/value*).

Subject A subject is either a *concept*, *entity*, *event*, or *property* (also collectively known as a '*resource*' in *RDF*), and is the item currently under consideration or focus; it is equivalent to the linguistic *subject*.

Subject extraction Subject extraction is an automatic process for retrieving and selecting subject names from existing *knowledge bases* or datasets. Extraction methods involve parsing and tokenization, and then generally the application of one or more information extraction techniques or algorithms.

SuperType SuperTypes (also Super Types) are a collection of (mostly) similar r*eference concepts*. Most of the SuperType classes have been designed to be (mostly) disjoint from the other SuperType classes. SuperTypes are synonymous with the *typologies* used in KBpedia. SuperTypes (typologies) provide a higher level of clustering and organization of *base concepts* for use in user interfaces and for reasoning purposes.

Supervised learning A *machine learning* task of inferring a function from labeled training data, which optimally consists of positive and negative *training* *sets*. The supervised learning algorithm analyzes the training data and produces an inferred function to determine the *class* labels for unseen instances correctly.

Symmetric Symmetric is when A relates to B exactly if it relates B with A.

Synechism Synechism is a philosophical doctrine that space, time, and law are continuous and form an essential *Thirdness* of reality in contrast to existing things and possibilities.

Tag A tag is a keyword or term associated with or assigned to a piece of information (*e.g.*, a picture, article, or video clip), thus describing the item and enabling keyword-based classification of information. Tags are usually chosen informally by either the creator or consumer of the item.

TBox A TBox (for terminological knowledge, the basis for T in TBox) is a 'terminological component,' that is, a conceptualization associated with a set of *facts*. TBox *statements* describe a conceptualization, a set of *concepts* and *properties* for these concepts. The TBox is sufficient to describe an *ontology*. Best practice often suggests keeping a split between instance records, the *ABox*, and the TBox schema.

Taxonomy In the context of knowledge systems, taxonomy is the hierarchical classification of *entities* of interest of an enterprise, organization, or administration, used to classify documents, digital assets, and other information. Taxonomies can cover virtually any type of physical or conceptual entities (products, processes, knowledge fields, human groups, etc.) at any level of granularity.

Thirdness Thirdness is one of the three universal categories and refers to all *classes*, *types*, or *generals*. The category conveys habitual law, continuity, or mediation.

Topic The topic (or theme) is the part of the proposition that is being talked about (predicated). In *topic maps*, the topic may represent any *concept*, from people, countries, and organizations to software modules, individual files, and events. Topics are in *Thirdness*.

Topic Map Topic maps are an ISO standard for the representation and interchange of knowledge. A topic map represents information using *topics*, associations (similar to a predicate relationship), and occurrences (which represent relationships between topics and information resources relevant to them), quite similar in concept to the *RDF triple*.

Training set A set of data used to discover potentially predictive relationships. In *supervised learning*, a positive training set provides data that meet the training objectives; a negative training set fails to meet the objectives.

Transitivity Transitivity is when item A is related to item B, and item B is related to item C, and then A is also related to A; this is the critical *property* for establishing inheritance chains.

Triple A basic *statement* in the *RDF* language, which is comprised of a *subject-property-object* construct, with the subject and property (and object optionally) referenced by *URIs*.

True negative A correct result, but one which fails (is negative) to meet the test objective. It is abbreviated TN.

True positive A correct result, and one which succeeds (is positive) to meet the test objective. It is abbreviated TP.

Tychism A philosophical doctrine that absolute chance is real and operative in the world; it is the source of irregularity and variety and the underlying force of evolution.

Type Types are the hierarchical classification of natural kinds of *instances* as determined by shared *attributes* (though not necessarily the same values for those attributes) and some common *essence*, which is the defining determinant of the type. All types may have hierarchy. Types are in Thirdness.

Typology Typologies are a natural organization of natural *classes* or *types*, with the most general types at the top of the hierarchy, and the more specific at the bottom. All types contained in a typology are children (subclasses) of the *root* type that is the basis for the character of the typology. Typologies provide a modular basis for expanding or collapsing the coverage of similar *instances* within a *knowledge base*. Typologies are central architectural components of KBpedia.

UMBEL UMBEL, short for Upper Mapping and Binding Exchange Layer, is an *upper ontology* of about 35,000 reference concepts, designed to provide common mapping points for relating different ontologies or schema to one another, and a vocabulary for aiding that ontology mapping, including expressions of likelihood relationships distinct from exact identity or equivalence. This vocabulary is also designed for interoperable *domain ontologies*.

Unsupervised learning A form of *machine learning*, this approach attempts to find meaningful, hidden patterns in unlabeled data.

Upper ontology An upper ontology (also known as a top-level ontology or foundation ontology) is an *ontology* that describes very general concepts that are the same across all knowledge domains. An important function of an upper ontology is to support very broad semantic interoperability between a large number of ontologies that are accessible ranking 'under' this upper ontology.

URI A uniform resource identifier is a Web-accessible address (string) for a specific piece of data; it is a more generalized form of a URL, which points to a page or resource location.

Value Value is a string, literal, or data value that pairs a numerical quantity, or quality or utility of a *subject* in relation to the meaning of its associated *attribute*, separate from the subject (but in association with it); these are known as key-value pairs; a value has no meaning or context absent its paired *attribute*.

Vocabulary A vocabulary, in the sense of knowledge systems or *ontologies*, is a *controlled vocabulary*. Vocabularies provide a way to organize knowledge for subsequent retrieval. They are used in formal declarations, subject indexing schemes, subject headings, thesauri, *taxonomies*, and other forms of knowledge organization systems.

Web-oriented architecture Web-oriented architecture, WOA, is a subset of the service-oriented architectural (SOA) style, wherein discrete functions are packaged into modular and shareable elements ('services') that are made available in a distributed and loosely coupled manner. WOA uses the representational state transfer (REST) style, geared to the HTTP model.

Wikidata This is a crowdsourced, open *knowledge base* of (currently) about 55 million structured *entity records*. Each record consists of *attributes* and values with robust cross-links to multiple languages. Wikidata is a key entities source.

Wikipedia Wikipedia is a crowdsourced, free-access, and free-content *knowledge base* of human knowledge. It has about five million articles in its English version. Across all Wikipedias, there are about 42 million articles in 288 different language versions.

WordNet WordNet is a lexical database for English that groups words into sets of synonyms called synsets; provides short, general definitions; and records the various semantic relations between these synonym sets. WordNet provides a combination of dictionary and thesaurus to support text analysis and artificial intelligence applications.

YAGO 'Yet another great ontology' is a WordNet structure placed on top of Wikipedia.

Index

M. K. Bergman, *A Knowledge Representation Practionary*,
https://doi.org/10.1007/978-3-319-98092-8