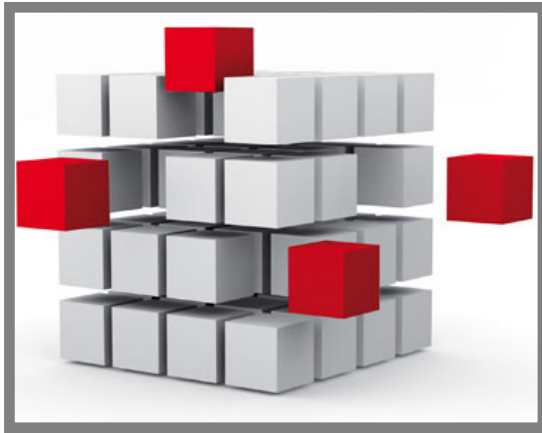


A Speculative Grammar for Knowledge Bases

by Mike Bergman - Monday, June 20, 2016

<http://www.mkbergman.com/1958/a-speculative-grammar-for-knowledge-bases/>



Finding a Natural Classification to Support AI and Machine Learning

Effective use of knowledge bases (KBs) for artificial intelligence (AI) would benefit from a definition and organization of KB concepts and relationships specific to those AI purposes. Like any language, the construction of logical statements within [KBAI](#) (knowledge-based artificial intelligence) requires basic primitives for how to express these arguments. Just as in human language where we split our words into roughly nouns and verbs and modifiers and conjunctions of the same, we need a similar primitive vocabulary and basic rules of statement construction to actually begin this process. In all language variants, these basic building blocks are known as the *grammar* of the language. A well-considered grammar is the first step to being able to construct meaningful and coherent statements about our knowledge bases. The context for how to construct this meaningful grammar needs to be viewed through the lens of the KB's purpose, which, in our specific case, is for artificial intelligence and machine learning.

In one of my recent major articles I discussed [Charles Sanders Peirce](#) and his ideas of the three universal categories and their relation to semiosis [\[1\]](#). We particularly focused on how Peirce approached categorization and its basis in his logic of [semiosis](#). I'd like to restrict and deepen that discussion a bit, now concentrating on what Peirce called the *speculative grammar*, the starting point and Firstness of his overall method.

The basic idea of the speculative grammar is simple. What are the vocabulary and relationships that may be involved in the understanding of the question or concept at hand? What is the "grammar" for the question at hand that may help guide how to increase our understanding of it? What are the concepts and terms and relationships that populate our domain of inquiry?

Hearing the term 'semiosis' in relation to Peirce most often brings to mind his [theory of signs](#). But for Peirce semiosis was a broader construct still, representing his overall theory of logic and truth-testing. Signs, symbols and representation were but the first part of this theory, the 'Firstness' or speculative

grammar about how to formulate and analyze logic.

Though he provides his own unique take on it, Peirce's idea of speculative grammar, which he erroneously ascribed to [Duns Scotus](#), can actually be traced back to the 1300s and the writings of [Thomas of Erfurt](#), one of the so-called Modists of the medieval philosophers [2]. Here is how Peirce in his own words placed speculative grammar in relation to his theory of logic [3]:

"All thought being performed by means of signs, logic may be regarded as the science of the general laws of signs. It has three branches: (1) *Speculative Grammar*, or the general theory of the nature and meanings of signs, whether they be icons, indices, or symbols; (2) *Critic*, which classifies arguments and determines the validity and degree of force of each kind; (3) *Methodeutic*, which studies the methods that ought to be pursued in the investigation, in the exposition, and in the application of truth." (CP 2:260)

Speculative grammar is thus a Firstness in Peirce's category structure, with logic methods being a Secondness and the process of logic inquiry, the methodeutic, being a Thirdness.

Still, What Exactly is a Speculative Grammar?

Charles S. Peirce's view of logic was that it was a formalization of signs, what he termed semiosis. As stated, three legs provide the basis of this formal logic. The first leg is a speculative grammar, in which one strives to capture the signs that most meaningfully and naturally describe the current domain of discourse. The second leg is the means of logical inference, be it deductive, inductive or abductive (hypothesis generating). The third leg is the method or process of inquiry, what came to be known from Peirce and others as [pragmaticism](#). The methods of research or science, including the scientific method, result from the application of this logic. The "pragmatic" part arises from how to select what is important and economically viable to investigate among multiple hypotheses.

In Peirce's universal categories, Firstness is meant to capture the potentialities of the domain at hand, the speculative grammar; Secondness is meant to capture the particular facts or real things of the domain at hand, the critic; and Thirdness is meant to capture methods for discovering the generalities, laws or emergents within the domain, the methodeutic. This mindset can really be applied to any topic, from signs themselves to logic and to science [1]. The "surprising fact" or new insight arising from Thirdness points to potentially new topics that may themselves become new targets for this logic of semiosis.

In its most general sense, Peirce describes this process or method of discovery and explication of new topics as follows [3]:

". . . introduce the monadic idea of »first« at the very outset. To get at the idea of a monad, and especially to make it an accurate and clear conception, it is necessary to begin with the idea of a triad and find the monad-idea involved in it. But this is only a scaffolding necessary during the process of constructing the conception. When the conception has been constructed, the scaffolding may be removed, and the monad-idea will be there in all its abstract perfection. According to the path here pursued from monad to triad, from monadic triads to triadic triads, etc., we do not progress by logical involution — we do not say the monad involves a dyad — but we pursue a path of evolution. That is to say, we say that to carry out and perfect the monad, we need next a dyad. This seems to be a vague method when stated in general terms; but in each case, it turns out that deep study of each conception in all its features brings a clear perception

that precisely a given next conception is called for." (CP 1.490)

The ideas of Firstness, Secondness and Thirdness in Peirce's universal categories are not intended to be either sequential or additive. Rather, each interacts with the others in a triadic whole. Each alone is needed, and each is irreducible.

As Peirce says in his *Logic of Relatives* paper [4]:

"The fundamental principles of formal logic are not properly axioms, but definitions and divisions; and the only facts which it contains relate to the identity of the conceptions resulting from those processes with certain familiar ones." (CP 3.149)

Without the right concepts, terminology, or bounding -- that is, the speculative grammar -- it is clearly impossible to properly understand or compose the objects or Secondness that populate the domain at hand. Without the right language and concepts to capture the connections and implications of the domain at hand -- again, part of its speculative grammar -- it is not possible to discover the generalities or the "surprising fact" or Thirdness of the domain.

The speculative grammar is thus needed to provide the right constructs for describing, analyzing, and reasoning over the given domain. Our logic and ability to understand the focus of our inquiry requires that we describe and characterize the domain of discourse in ways that are properly scoped and related. How well we bound, characterize and signify our problem domains -- that is, the speculative grammar -- directly relates to how well we can reason and inquire over that space. It very much matters how we describe, relate and define what we analyze and manipulate.

Let's take a couple of examples to illustrate this. First, imagine [van Leeuwenhoek](#) first discovering "[animacules](#)" under his early, advanced microscopes. New terms and concepts like flagella, cells, and vacuoles needed to be coined and systematized in order for further advances in microorganisms to be described. Or, second, imagine "[action at a distance](#)" phenomena such as magnetic repulsion or static electricity causing hair to stand on end. For centuries these phenomena were assumed to be caused by atomistic particles too small to see or discover. Only when [Hertz](#) was able to prove [Maxwell's equations](#) of electromagnetism hundreds of years later in the mid-1800s were the concepts and vocabulary of waves and fields sufficiently developed to begin to unravel electromagnetic theory in earnest. Progress required the right concepts and terminology.

For Peirce, the triadic nature of the sign -- and its relation between the sign, its object and its interpretant -- was the speculative grammar breakthrough that then allowed him to better describe the process of signmaking and its role in the logic of inquiry and truth-testing (semiosis). Because he recognized it in his own work, Peirce understood a conceptual "grammar" appropriate to the inquiry at hand is essential to further discovery and validation.

How Might a *Speculative Grammar* Apply to Knowledge Bases?

Perhaps one way to understand what is intended by a speculative grammar is to define one. In this instance, let's aim high and posit a grammar for knowledge bases.

Since we had been moving steadily to a typology design for our entities and were looking at all other aspects of structure and organization of the knowledge base (KB) [5], we decided to apply this idea of a speculative grammar to the quest. We consciously chose to do this from two perspectives. First, in keeping with Peirce's sign trichotomy, we wanted to keep the interpretant of an agent doing artificial intelligence front-and-center. This meant that the evaluative lens we wanted to apply to how we conceptualized, organized and provided a vocabulary for the knowledge space was to be done from the viewpoint of machine learning and its requirements. Once we posit the intelligent agent as the interpretant, the importance of a rich vocabulary and text (NLP), a well-formed structure and hierarchy (logical inference), and a rich feature set (structure and coherent characterizations), becomes clear.

Second, we wanted to look at the basis for organizing the KB concepts into the Peircean mindset of Firstness (potentials), Secondness (particulars) and Thirdness (generals). Our hypothesis was that conforming to Peirce's trichotomous splits would help guide us in deciding the myriad possibilities of how to arrange and structure a knowledge base.

We looked as well as to what language to write these specifications. Our current languages of [RDF](#), [SKOS](#) and [OWL](#) could capture all first-order logic imperatives, but the OWL annotation, object and datatype properties did not exactly conform to the splits we saw [11]. On the other hand, tooling such as [Protégé](#) and the [OWL API](#) were immensely helpful, and there are quite a few supporting tools. Formalisms like [conceptual graphs](#) were richer and handled higher-order logics, but also lacked tooling and widespread use. Since we knew we could adapt to OWL, we stuck with our original language set.

Guarino, in some of the earliest (1992) writings leading to semantic technologies, had posited knowledge bases split into concepts, attributes and relations [6]. This was close to our thinking, and provided comfort that such splits were also being considered from the earliest days of the semantic Web. Besides Peirce, we studied many philosophers across history regarding fundamental concepts in knowledge organization. [Aristotle's categories](#) were influential, and have mostly stood the test of time and figured prominently in our thinking. We also reviewed efforts such as Sarbo's to apply Peirce to knowledge bases [7], as well as most other approaches claiming Peirce in relation to KBs that we could discover [8-10].

Because the intent was to create a feature-rich logic machine for AI, we of course wanted the resulting system to be [coherent](#), [sound](#), [consistent](#), and relatively [complete](#). Though the intended interpretant is artificial agents, training them and vetting results still must be overseen by humans in order to ensure quality. We thus wanted all features and structural aspects to be described and labeled sufficiently to be understood and interpreted by humans. These aspects further had to be suitable for direct translation into any human language. For interchange purposes, we try to use [canonical forms](#).

Once these preliminaries are out of the way, the task at hand can finally focus on the fundamental divisions within the knowledge base. In accordance with the Peircean categories, we saw these splits:

- Firstness -- these 'potentials' include base concepts, attributes, and relations in the abstract
- Secondness -- these are the 'particular' things in the domain, including entities, events and activities
- Thirdness -- the 'generals' in the knowledge base include classes, types, topics and processes.

Ultimately, features richness was felt to be of overriding importance, with features explicitly understood

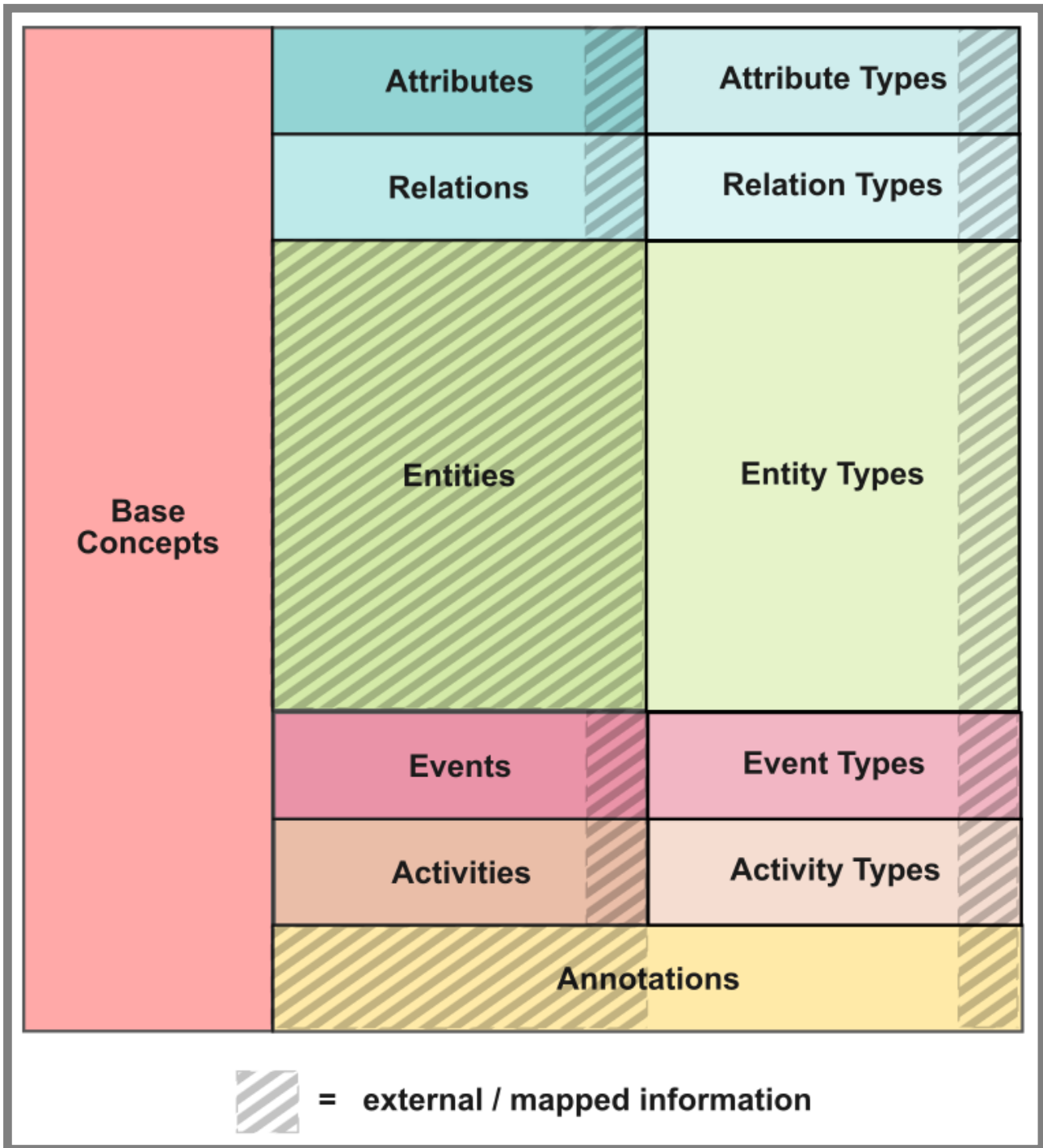
to include structure and text.

A Speculative Grammar for Knowledge Bases

With these considerations in mind, we are now able to define the basic vocabulary of our knowledge base, one of the first components of the speculative grammar. This base vocabulary is:

- **Attributes** are the ways to characterize the entities or things within the knowledge base; while the attribute values and options may be quite complex, the relationship is monadic to the subject at hand. These are intensional properties of the subject
- **Relations** are the way we describe connections between two or more things; relations are external-facing, between the subject and another entity or concept; relations set the extensional structure of the knowledge graph [\[11\]](#)
- **Entities** are the basic, real things in our domain of interest; they are nameable things or ideas that have identity, are defined in some manner, can be referenced, and should be related to types; entities are the bulk of the overall knowledge base
- **Events** are nameable sequences of time, are described in some manner, can be referenced, and may be related to other time sequences or types
- **Activities** are sustained actions over durations of time; activities may be organized into natural classes
- **Types** are the hierarchical classification of natural kinds within all of the terms above
- The **Typology** structure is not only a natural organization of natural classes, but it enables flexible interaction points with inferencing across its 'accordion-like' design (see further [\[5\]](#))
- **Base concepts** are the vocabulary to the grammar and top-level concepts in the knowledge graph, organized according to Peircean-informed categories
- **Annotations** are indexes and the metadata of the KB; these can not be inferenced over. But, they can be searched and language features can be processed in other ways.

How these vocabulary terms relate to one another and the overall knowledge base is shown by this diagram:



A Knowledge Base Grammar

As part of the ongoing simplification of the [TBox \[12\]](#), we need to be able to distinguish and rationalize the various typologies used in the system: attributes, relations, entities, events and activities. Here are some starting rules:

- **RULE:** Entities can not be topics or types
- **RULE:** Entities are not data types; these are handled under values processing
- **RULE:** Events are like entities, except they have a discrete time beginning and end; they may be nameable
- **RULE:** Activities (or actions) act upon entities, but do not require a discrete time beginning or end
- **RULE:** Attributes are not metadata; they are characteristics or descriptors of an entity
- **RULE:** Topics or base concepts do not have attributes
- **RULE:** Entity types have the attributes of all type members
- **RULE:** Relation types do not have attributes (also, relations do not have attributes)
- **RULE:** Attribute types do not have attributes (also, attributes do not have attributes)
- **RULE:** All types may have hierarchy
- **RULE:** No attributes are provided on relations (as in E-R modeling), just annotations
- **RULE:** Annotations are not typed, and can not be inferred over.

As we work further with this structure, we will continue to add to and refine these governing rules.

The columns in the figure above also roughly correspond to Peirce's three universal categories. The first column and part of the second (attributes and relations) correspond to Firstness; the remainder of the second column corresponds to Secondness; and the third column corresponds to Thirdness. I'll discuss these distinctions further in a later article.

In combination, this vocabulary and rules set, as allocated to Peirce's categories, constitutes the current speculative grammar for our knowledge bases.

Recall that the interpretants for this design are artificial agents. It is unclear how the resulting structure will be embraced by humans, since we were not the guiding interpretant. But, like being able to readily discern whether an object is plumb or level, humans have the ability to recognize adaptive structure. I think what we are building here will therefore withstand scrutiny and be useful to all intelligent agents, artificial or human.

Conclusion

Generating new ideas and testing the truth of them is a logical process that can be formalized. Critical to this process is the proper bounding, definition and vocabulary upon which to conduct the inquiries. As Charles Peirce argued, the potentials central to the inquiries for a given topic need to be expressed

through a suitable *speculative grammar* to make these inquiries productive. How we think about, organize and define our problem spaces is central to that process.

The guiding lens for how we do this thinking comes from the purpose or nature of the inquiries at hand. In the case of machine learning applied to knowledge bases, this lens, I have argued, needs to be grounded in Peirce's categories of Firstness, Secondness and Thirdness, all geared to feature generation upon which machine learners may operate. The structure of the system should also be geared to enable (relatively quick and cheap) creation of positive and negative training sets upon which to train the learners. In the end, the nature of how to structure and define knowledge bases depends upon the uses we intend them to fulfill.

[1] M.K. Bergman, 2016. "[A Foundational Mindset: Firstness, Secondness, Thirdness.](#)" *AI3::Adaptive Information* blog, March 21, 2016.

[2] Alessandro Isnenghi, 2008. "A Semiótica de CS Peirce e a Gramática Especulativa de Modistae" (or, "[C.S. Peirce's Semiotic And Modistae's Grammatica Speculativa](#)"), *Cognitio-Estudios: revista eletrônica de filosofia*. ISSN 1809-8428 5, no. 2 (2008).

[3] See the [electronic edition](#) of *The Collected Papers of Charles Sanders Peirce*, reproducing Vols. I-VI, Charles Hartshorne and Paul Weiss, eds., 1931-1935, Harvard University Press, Cambridge, Mass., and Arthur W. Burks, ed., 1958, Vols. VII-VIII, Harvard University Press, Cambridge, Mass. The citation scheme is volume number using Arabic numerals followed by section number from the collected papers, shown as, for example, CP 1.208.

[4] This quote is drawn from Charles Sanders Peirce, 1870. "Description of a Notation for the Logic of Relatives, Resulting from an Amplification of the Conceptions of Boole's Calculus of Logic", *Memoirs of the American Academy of Arts and Sciences* 9 (1870), 317-378 (the "Logic of Relatives"), using the same numbering as from [3].

[5] M.K. Bergman, 2016. "[Rationales for Typology Designs in Knowledge Bases.](#)" *AI3::Adaptive Information* blog, June 6, 2016.

[6] Nicola Guarino, 1992. "[Concepts, Attributes and Arbitrary Relations: Some Linguistic and Ontological Criteria for Structuring Knowledge Bases.](#)" *Data & Knowledge Engineering* 8, no. 3 (1992): 249-261. Continuing in the same vein, see also Nicola Guarino, 1997. "[Some Organizing Principles for a Unified Top-level Ontology.](#)" in *AAAI Spring Symposium on Ontological Engineering*, pp. 57-63. 1997. Also, for a general view of ontology at that time, see Thomas R. Gruber, 1993. "[A Translation Approach to Portable Ontology Specifications.](#)" *Knowledge Acquisition* 5, no. 2 (1993): 199-220.

[7] Auke JJ Van Breemen and Janos J. Sarbo, 2009. "[The machine in the ghost: The syntax of mind.](#)" *Signs-International Journal of Semiotics* 3 (2009): 135-184. and Janos J. Sarbo and József I. Farkas, 2002. "[On the isomorphism of signs, logic and language.](#)"

[8] Lehmann, Fritz, and Rudolf Wille. "[A triadic approach to formal concept analysis.](#)". Springer Berlin Heidelberg, 1995.

[9] József István Farkas, 2008. "[A Semiotically Oriented Cognitive Model of Knowledge Representation.](#)" Ph.D. thesis, *Radboud University of Nijmegen*, April 23, 2008.

[10] John F. Sowa, 1995. "Top-level Ontological Categories," *International Journal of Human-computer Studies* 43, no. 5 (1995): 669-685.

[11] Attributes, Relations and Annotations comprise OWL properties. In general, Attributes correspond to the OWL datatypes property; Relations to the OWL object property; and Annotations to the OWL annotation property. These specific OWL terms are not used in our speculative grammar, however, because some attributes may be drawn from controlled vocabularies, such as colors or shapes, that can be represented as one of a list of attribute choices. In these cases, such attributes are defined as object properties. Nonetheless, the mappings of our speculative grammar to existing OWL properties is quite close.

[12] As [earlier wrote](#), "Description logics and their semantics traditionally split concepts and their relationships from the different treatment of instances and their attributes and roles, expressed as fact assertions. The concept split is known as the TBox (for *terminological* knowledge, the basis for T in 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 (for assertions, the basis for A in 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." In the diagram above, the middle column represents particulars, or the ABox components. The definition of items and the first and third columns represent TBox components.

PDF generated by *AI3::Adaptive Information* blog