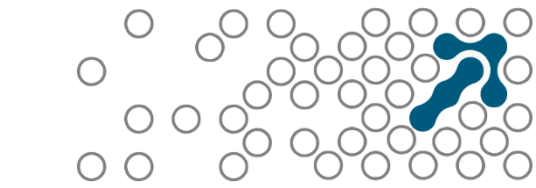
VERACITY

By: Dave McComb



semantic arts

Encarta defines veracity as "the truth, accuracy, or precision of something" and that seems like a pretty good place to start.

Our systems don't model uncertainty very well, and yet that is exactly what we deal with on a day-to-day basis. This paper examines one aspect of modeling certainty, namely veracity, and begins a dialog on how to represent it.

Veracity

Encarta defines veracity as "the truth, accuracy or precision of something" and that seems like a pretty good place to start.

In our case we will primarily be dealing with whether a symbolic representation of something in the real world faithfully represents the item in the real world. Primarily we are dealing with these main artifacts of systems:

- Measurements – is the measurement recorded in the system an accurate reflection of what it was meant to measure in the real world?
- Events – do the events recorded in the system accurately record what really happened?
- Relationships – do the relationships as represented in the system accurately reflect the state of affairs in the world?
- Categorization – are the categories that we have assigned things to useful and defensible?
- Cause – do our implied notions of causality really bear out in the world? (This also includes predictions and hypotheses.)

Only the first has ever received systematic attention. Fuzzy numbers are a way of representing uncertainty in measurements, as is "interval math" and the uncertainty calculations used in Chemistry (2.034 +/- .005 for instance).

But in business systems, all of these are recorded as if we are certain of them, and then as events unfold, we eventually may decide not only that we are not certain, but that we are certain of an opposite conclusion. We record an event as if it occurred and until we have proof that it didn't, we believe that it did.

In this paper we're going to discuss a bit where our ideas of certainty come from, we'll leave it to subsequent articles to apply this to systems.

Certainty

How certain are we of events in the world? Generally we proceed as if we know what has happened or what is going to happen, but in some circumstances these certainty gets violated. Let's look at some everyday examples:

Category	**Circumstances of failure**	**Odds of Certainty being violated**
Gravity: certainty that a dropped object that seemed to be massive will fall	Magic, objects that appear massive but are lighter than air, in a field of magnetism	1;1,000,000 We very rarely see this
Inertia: a moving object (say a fly ball to left field) will continue its trajectory	Interception from unseen object, explosion, magic	1:1,000,000
Persistance: a physical object will still exist from one minute to the next	Explosion, magic	1:1,000,000
Mechanical failure: elevators or airplanes	Plane crash or elevator crash	1:100,000 this is probably low, there are only a few plane crashes per year and millions of flights
Mechanical failure: Car	Car breaks down, varies depending on age and reliability	1:1000 – 1:10,000 (if we make 1500 trips a year, we would expect to get the car to fail at least every 7 years, and maybe as much as annually)
Mechanical failure: cell phone	How often you were in what you believed to be cell phone territory and could not send or did not receive a call	1:20 ("can you hear me now?")
Security: "have" a security token that is not yours	How often will someone have something (like a dongle or a key) that doesn't "belong" to them	1:100 (get statistics on identity theft plus some other stuff)

Security: "are" biological markers that are mistaken	How often are fingerprints or iris scans "wrong"	1:100,000 (these have a pretty low false positive rate, by the way they aren't of much use for identifying, merely confirming)
Security: "know" you know something that only you should know	How often does someone else know your password, or mothers maiden name or even SSN or Claim number	1:20 (not that it creates a problem that often, but it is fairly common for one person to use anothers access to get onto a system)
Security: "do" what can you do that verifies you are who you say you are	Being able to respond to a piece of email or postal mail directed at a person is a proxy for that persons identity	1:100,000 (be interesting to get some statistics, see the EBIA article)
Security: personal facial recognition error	You see somebody and mistake them for someone else	1:10,000 (we do a pretty good job of this but still manage maybe one per year)
Property misidentification	We rely on attributes of property to identify it. How often do we, for instance, misidentify our car in a parking lot or our suitcase on a luggage carosel	1:1,000 (you would think this would be based on how common the attributes are, black wheely suitcases for instance, but people with the black wheelies know there's looks like many others and take pains to distinguish theirs)

Numerical identifiers (so a phone number for a person for instance, or a part number for a part)	How often do we dial a wrong number of type in a part number or a customer number and get the wrong one?	1:100- 1:500 (the error rate goes up drastically as the number of digits exceeds 7, with the number of transcriptions and the half life of the data – phone numbers will soon start getting better with number portability)
Mail will get delivered	Odds that a piece of mail addressed to someone will get to them	1:50 (mailers with incentives and frequent mailing are only able to achieve 98% deliverability, which is not a high as getting to the right person, many lists are much worse than this.)
Web site available	Odds that an url is reachable	1:10 (Even google who do a good job of keeping this stuff up to date, deals out urls that are unreachable quite frequently.)

This is quite a range of certainty. And we've only touched some of the categories we mentioned under the veracity heading. One of the things that we can notice from the above list is that the range of certainty is many orders of magnitude.

Veracity Measure First Cut

So a first cut veracity measure would be a logarithmic scale, much like the Richter scale for earthquakes or the Decibel scale for sound.

9	Well established law of physics, like gravity.
8	Almost infallible. DNA identification
7	Reliable, fallible, but incredibly reliable. Odds of a commercial airline making a successful flight
6	Highly reliable, recognition of a person or property, reliability that a late model car will start.
5	Rely on confidently. Odds that a check with proper ID will not bounce
4	Rely on. Mistakes happen occasionally, such as wrong numbers.
3	Rely on, but not with a lot of confidence. Odds that our cell phone will work or that mail will get to someone, or that a flight will be on time.
2	Better than random. Odds that a person named Dave is a male
1	No better than random chance

What I want to do in this paper is give us a language with which to talk about veracity. The implications are left to many more papers yet to be written, and discussions yet to be had.

11 Old Town Square
Suite 200
Fort Collins, CO 80524

970-490-2224
305-425-2224
info@semanticarts.com