ENTROPY IN INFORMATION THEORY: A PARADOX

*David Batty**

As we have heard direct analogy of entropy implies a progression from low entropy (everything in order, waiting to be used) to high entropy (everything mixied up or impossible to find to be used). Campbell's (Grammatical Man) analogies refer to a library after use by a class of high school kids on a rush project.

Campbell also sees the librarian as Maxwell's demon–reconverting the random and chaotic collection to order. The proof that Maxwell's demon cannot exist depends on the proposition that the demon creates as much entropy as he eliminantes, because he uses other information to organize the information he is working on. "To be precise, information that is not really in his possession because he guesses it." (Leo Szilard)

The original analogy of the library in chaos is not the best. True–but expected–and the librarian MUST continually reorganize.

But that reorganization is done:

a) as a predicted expense
b) the expense is lessened–according to predefined rules known to (even devised by) the librarian

The analogy is weakened further by the assumption that the library starts out in order. Random order of accession.

---

* C. David Batty. President, C.D.B. Enterprises.

This realization brings us closer to a different application of the entropy image (as in poetic image).

The information scientist is indeed a kind of Maxwell's demon—bringing order into chaos. But it is not the RE-creation of order. The chaos is the entirely random world of primary source information produced at need/will by scientists, authors, etc.—primal chaos. The order is the introduction of bibliographic control according to the techniques, etc. of the information scientist.

However, notice that the CHAOS is not any longer the HIGH entropy of expanded gases in a heat driven engine, where previously easily distinguishable molecules of low entropy have now become an indistinguishable (and undistinguished) mass. The chaos is rather the as yet undistinguishable mass of unorganized literature. In other words, we have to stand the entropy image on its head, and BEGIN with high entropy, and then proceed to UNMIX the already mixed up.

By applying classification and indexing techniques, by building well organized and easily accessed databases of comprehensible, brief descriptions and encapsulations, the information scientist begins to create the state of LOWER entropy in which information is labelled, compartmented or made available for retrieval.

Campbell notes approvingly that the state of lowest entropy in his chaotic library would be one in which every item had only one place, and could be found in it. Sounds like the kind of ambition Spofford and Putnam had for LC. Or the Harvard librarian crossing the Yard.

Actually such an extreme state of low entropy is BAD. First, it costs too much in prior effort. Also it assumes a simplistic view of the information explosion. (expand)

Second, it does not allow or support across-fertilization of ideas. The original high entropy state is TOO chaotic for all but the polymath. (expand)

The processes of analysis and (more important) synthesis

in information science tools reduces the level of entropy to manageable proportions. But must be able to accommodate different modes or directions of organization at need/will.

One aspect of the Maxwell demon analogy and its refutation is certainly true:

a) information outside the information being handled is necessary
b) there is an equation of effort in all such activity whose sides must balance

That is, there is a necessary expenditure of energy in communicating the information. We may move it to different parts of the equation—we may even share it out among different functions in the equation—but we cannot get rid of it.

At one extreme, if we do NOTHING to the primal chaos of raw primary information, we remove the burden of effort from our shoulders (indeed we never accept it in the first place) and leave it squarely on the shoulders of the users of the information who will have to hunt at random (or nearly at random) across the literature of their immediate discipline and probably across other disciplines as well.

Remember the dimensions of information:

- OF TIME: that information has always generated in the past for past needs, but will be used in the future for needs as yet unspecified
- OF SPACE: that information is generated at the bottom end of vertical hierarchies of increasing specificity—but will be used across lateral discipline boundaries.

At the other extreme of activity we can invent a complete, monolithic, enumerative classification of everything and fit all knowledge into it—each item in its place—a kind of universal straitjacket. Here the information scientist has taken on ALL the effort. (But again, remember the dimension of information.)

Clearly the answer (as I implied earlier) lies in the middle

ground. SOME organization but NOT TOO MUCH. This also shares the effort between the information scientist and the user.

Sometimes the effort by the user is sweat and sometimes comprehension.

KWIC/citation indexing

faceted organization

The effort does not have to be divided between organizing and using. The user can help in the organizing–

SUBTIL

Have talked in terms of classification etc. and ISAR. But also works for all other processes of bibliographical control.

Bibliographies of special subjects

Bibliographies over several subjects

Cross disciplinary/mission oriented bibliographies

If some organization is good–to reduce high entropy to low, then is not more organization better?

No. Too much organization EITHER produces the universal straitjacket–that the user has to fight (expend effort) to get out of and look over boundaries.

(Creation and the bisociative process – Koestler – expand)

OR produces an over-homogenized, generalized soup. (Mix all the cans of soup at once. . .)

Entropy has been described as the reduction of energy from an accessible state to an inaccessible state. Information in its primary form is not easily accessible (you have to read all of it), but take application of order reduces entropy to bring information to an ordered state of accessibility. Too much reduction brings it to the level at which the information is again inaccessible (because concealed or ignored behind the bland exterior of generalization).

Information science has to be the guardian of the entropy reducing process–to start it up and also to stop it before it takes over. (Before Maxwell's demon begins to eat into the capital instead of the interest–i.e., the actual REDUCTION of the value of information by escalating cost of overhead.)

Paradoxically the extreme of low entropy is as inhibiting as high entropy was, and for much the same reason—that the user of the information must become his own Maxwell's demon—and that reduces his productivity.

Productivity depends on the middle ground or phase of the entropy process.

Theoretically we can discuss entropy in terms of information organization and disorganization, or in terms of an on-going process. But in the end we must treat it as a phenomenon that can be brought under our control and put to practical, user-oriented ends of increased productivity of and through information.