tripleC 4(2): 172-177, 2006
ISSN 1726-670X
http://tripleC.uti.at

Available information — preparatory note for a theory of information space

Lars-Erik Janlert

The Department of Computing Science,
Umeå University, Umeå, Sweden
lej@cs.umu.se

Abstract: Information technology warps information space, but there are limits to the availability of information. *Information distance* is introduced to begin investigate the shape of information space, which is very much needed. The concept of *availability profile* is proposed as a way of defining spatial location in information space, also interpretable as information state. A first check of the possibilities to extend the agentcentric view into an infocentric view is not immediately discouraging, but many problems and issues remain.

Keywords: information space, information distance, availability profile

Acknowledgement: Thanks to Ingvar Johansson for comments on an earlier version.

1 Information space is warped by information technology

We can imagine how in the distant past the information space of sentient beings more or less coincided with physical space: to learn some contingent fact of the world you would have to be there and at that particular moment and actually sense it. In the age of information, the shape of information space is brought to deviate more and more from the shape of physical space. Establishing a telegraph or phone connection brings objects distant in physical space close in information space. Books draw together information scattered in physical space, making it readily available at each point in physical space where a copy is at hand. The newest information technologies bring this development to a head. Information space seems to be contracting and the current technological visions of making all information instantly available everywhere may make us wrongly believe that information space will eventually collapse into a single, omnipresent, point.

The term *information space* I am using here is meant to stand for a strict and specific concept; the sharp version of a rather common but vague and unreflecting notion. This common notion is present in the widely used but problematic spatial metaphor for information, knowledge and communication. It is a perspective revealed in common expressions such as “flow of information,” “information storage and retrieval,” “information landscape,” etc., taking information to be a commodity which is located and can be transported from one point to another. It is also a commonplace that movement in physical space—for example, a walk from the coffee room to your office—entails movement also in information space: some information sources getting closer (your personal files, selection of reference books, etc.), others becoming more distant (your colleagues, postings on the bulletin board, etc.). We have grown accustomed

to thinking of information in terms of space, but our uncertainty about where information really is also grows. What kind of space does information exist in, what is its shape and limits, and what relation does it have to physical space?

2 Limits of availability

That there are limits to how much information can be crammed into the information neighborhood of an agent (information user) is easily illustrated.

Consider products of integers. A compact little algorithm can produce any requested product of two integers. Inevitably, when larger numbers are multiplied, it will take longer time to produce the result. An alternative would be an enormous multiplication table that explicitly records products of two numbers. If you need to know a certain product, just look it up. Inevitably, some products will take longer time to look up than others, because the products, the physical inscriptions, will take up physical space and hence have varying physical distances to the origin of the request. The line of reasoning here is based on rather well established principles: (a) computation takes time, and the bigger the task, the more time; (b) information needs a physical carrier which has some minimal physical extension, so in a limited volume, there can only be a limited amount of information; (c) there is an upper limit to signaling speed. (At this time, let us leave quantum theory out of the picture, setting a more modest but still daunting goal of attempting to understand "classical information space.")

As a variation, imagine that your job requires you to find products of integers and that your only external support for this task is a desktop and a small piece of paper on which is written "3177 · 11433 = 36322641." At some point you need to know the product of 3177 and 11433; you pick up the slip of paper, see that the numbers match, and read off the result. (There is only one piece of paper, so you are very lucky.) A slight change of scene: now imagine millions of tiny paper slips recording various integer products, all at arm's length. (Notice the blurred line between "having" and "attending to" from the availability point of view.) It will take some time to locate the right piece of paper. In this situation we could find ways of organizing the papers and use certain procedures (algorithms) to speed up search. We must realize, however that any such device will also add to the total time and/or space requirements, in competition with the "raw data" themselves.

There are many examples of practical manifestations of these limitations to availability; one particularly well-known is the layered arrangement of different kinds of memories in a computer. Very fast but small memories closest to the central processing, slower but more spacious memories as we proceed outwards, to so-called "external" "mass" memories. You might think that the reason that the very fastest memory technologies are not used to cover the entire requirements of memory is the cost. Yes, but I suggest that the high cost can be seen as an expression of the fact that there is limited information space close to the central processor; the more restricted the closer you get; real estate prices soar as we approach the center precisely because real estate is limited.

Examples such as these strongly suggest that the goal of making all information equally and instantly at hand (not to mention everywhere) is not just very hard to attain in practice, but is not even a very useful ideal: it is *impossible in principle*.

3 The importance of information space exploration

Information society shifts our focus of attention to information, knowledge and idea transfer, to idea production, to idea products in a wide sense, ranging from scientific and technological knowledge to media and entertainment. Industrial society focused on material products and material production processes; information society is focused on information and ideas and information and thought processes. When people make their living on assembling, assessing, designing and creating information, and society as a whole has become critically dependent on fast, abundant, timely and relevant information for its workings, there is a mounting pressure to arrange and organize information to meet our needs in the most efficient way.

Information overload is one example of a new phenomenon signaling that all is not well with the current knowledge regime (Toffler 1970). *Context-Aware Computing* is one example of many endeavors to improve the situation (Moran 2001). An example of a common strategy to focus relevant information builds on the idea that information that has some bearing on the physical situation we are in, is generally more likely to be of use than information that doesn't. At work in this case is the (hidden) assumption of a *proximity principle* of contextual relevance: what matters is close, and that which is close matters. But, again, technological advancements tend to compromise the topology of closeness.

In times as these, when information space seems to be constantly bending and twisting in ever new ways, in a society where information has become central to living, it would seem to be a very good idea to begin the exploration of information space, in order to understand it better and possibly improve the situation.

4 The prospects of information space research

What, then, are the prospects for making sense of "information space" as literally some kind of space (and not just a nice metaphor)? There is certainly no guarantee of success and it is not obvious how far we can get.

Speaking for the project of space exploration is firstly the connection to and suggested original alignment with physical space. One way of thinking of information space is that it has a subspace which is (isomorphic to) physical space: the physical dimensions are represented plus a non-physical extension of unknown dimensions that have come to play a more and more important role, which we experience informationally as a warping of physical space. Secondly, there are the theoretically as well as empirically established limitations to the availability of information, which frequently are formulated in terms of some *tradeoff*, as between time and memory requirements. Third, it is a fact that we are often quite successful at understanding things in terms of some more or less abstract space (it helps, of course, that "space" is such a wide and elastic notion).

So let us bravely aim for the theoretically most attractive kind of space, a metric space. Even if this might be to set hopes too high, some of the preparatory work may still prove useful.

5 Virtual availability and information distance

In order to explore the shape and laws of information space, first of all a concept of distance is needed, something to serve as a measuring stick. What we need is a measure that robustly abstracts from details of how information is implemented.

To ascertain the current phase of the moon for example, you might use methods as different as looking up at the sky, consulting your pocket calendar, applying some little algorithm to the current date, or calling your hobby-astronomer friend. In the computerized and networked world of today, you often cannot say and neither do you care whether information is locally stored, or produced on demand by sensors or by calculations, or brought to you from some remote database, or delivered by any combination of these and possibly other methods: the result is virtually the same. Indeed, the very same information request may be satisfied in different ways on different occasions. It is this *virtual availability* of information, information distance should capture.

I propose *information distance* as a basic conceptual tool, a necessary prerequisite, for beginning to explore information space. The distance to a piece of information is the time it takes to access it, to bring it to the fore, to produce it. The unit of measure is obvious and simple: seconds. The exact *definition* of distance and the proper *conditions* and *operations* of measuring are less obvious, however.

6 To have and have not – dichotomy abandoned

The notion of information distance means that we abandon the traditional strict *have – have-not* dichotomy of information (and knowledge) for a more precise and graded measure of availability. Note that this is not some kind of fuzzy theory of information; it does not have to do with degree of certainty,

applicability or probability. It is a matter of being able to *produce in x seconds*, for different values of *x*. (The basic idea was put forward in Janlert 1985.)

Contrast this model of information availability with the traditional, formalistic model of having information (knowledge), in which the information (knowledge) you have is modeled as a set of propositions closed under logical inference. On the one hand it does not make a distinction between "immediately available" information ("axioms," say) and information only derivable via arbitrarily long and tortuous chains of inferences. If you know Peano's axiom for the natural numbers, then you also know that $3714089895285 \cdot 2^{60000} - 1$ is a prime number (because it can be proved). On the other hand it creates a sharp, impenetrable border to the world outside, whereas it is reasonable to assume that "real agents" have a rather open-ended possibility of acquiring information, by using any means available in a world too rich to exhaustively capture in any finitely defined set of propositions, and overflowing with various shortcuts to information.

Generally, the utility of (some particular) information will decrease monotonically with its distance. In some cases, there will be a rapid drop and a rather sharp boundary beyond which the information suddenly ceases to be of interest (for instance, because an opportunity window has closed). In such a case, in a sense, we do not "have" the information (in time). Obviously this is situation dependent, and the analysis and discussion of such issues should be helped by a more abstract and absolute notion of distance.

The shift from a strict dichotomy of possession to a gradual and boundless availability metrics gives rise to a number of questions. For example, if you have information P, within some "reasonable" limit L, say, how far away is the information that you have it? If having the information that P is available means that you have the information that P in full detail (referential transparency), then, clearly, it is (almost) tantamount to having P itself, and they are thus equidistant. If having the information that P is available is interpreted in the weaker sense that we have some kind of indication of P's availability—within L, or with some more precise measure—then that information *can* be closer, but it does not have to be.

Example. Take "$3177 \cdot 11433 = 36322641$" as P. The information that "$3177 \cdot 11433 = 36322641$" is available is at the same distance as P itself. The information that the product of 3177 and 11433 is available *may* be closer. The information that all products of integers less than 100000 are available *may* be closer. We might also have inconclusive information, for example that *some* products of integers are available (within L). The further away some information is expected (feared) to be, the more interesting indicators and predictors of its distance will become. How much time do we afford to spend in pursuing information that might be beyond L, or, generally more distant than we surmised?

7 Availability profile

Given that we have a measure of how close or distant a certain piece of information is, its information distance, we can consider the whole range of distances to various pieces of information, call it *the availability profile*. We can think of it as a mapping from information to distance (i.e. time; unavailable information can be mapped to □). The availability profile may be used to define an agent's location in information space, and it may be interpreted as the agent's information state. Movement in physical space will normally change the availability profile, moving or changing objects in physical space will also typically change the availability profile. As a consequence, this interpretation implies that the environment takes part in the information state of the agent.

8 Agentcentric view vs. infocentric view

So far, these considerations assume an agentcentric point of view: how distant information is for a particular "agent," or information user. Can the agentcentric view be expanded into an infocentric view, taking information *itself* to be in some sense located in information space, in a way that makes space and distances coherent? *Where* is the information?

There hardly seems to be any other option than to say that it is located at any point where the distance to it is (approaches) zero. For example, suppose the agent is located at some point A in information space (which we tentatively take to be defined by the availability profile, the total information state). Were the agent to be at another point B a certain piece of information X could be accessed in time ∂ (B, X). We could now argue that the triangular inequality

∂ (A, X) $\leq \partial$ (A, B) + ∂ (B, X)

plausibly holds—given that ∂ (A, B) is interpreted as the time it takes to move from point A to point B in information space. It should not take longer to get from information state A to state X, than by first getting to B, and from there getting to X, because that is one way of doing it.

But I have already in the beginning suggested that the same almost zero-time-accessible piece of information can occur at several points in physical space (copies of the same book, e.g.). Assuming that a coherent view is indeed possible, from this we may conclude that pieces of information are generally not point-like, but rather some kind of extended manifolds that may appear disconnected in physical space but hopefully are connected in the higher-dimensional information space. (Generally, it is implausible that these different points in physical space are actually the same point in information space. Does it still remain a logical possibility, as in two "totally" communicating points in physical space? Would that not imply that the two points were also physically identical, albeit some convoluted physical topology might be involved, the intersection point of two distinct physical contexts?)

It is apparent that this line of reasoning gets rather involved in using the poorly defined concept of "piece of information"? What is that? How does granularity and interdependencies get into the picture (or not)? I will return shortly to the "piece of information" issue. At least as problematic is the involvement of an "agent."

9 Thick and thin agents

What is the cognitive equipment of an agent? A thick agent carries around information in the physical world, inside its body (memories, cognitive abilities), or outside (notebooks, palm pilots). By moving physically it not just changes its physical distances to different information sources, it also by its own presence (and possible actions) rearranges the information sources and instruments of the physical world (with consequences for itself and for other agents).

A problem with thick agents is a potential lack of intersubjectivity—information distances may come to depend not only on where you are in information space, but also on who you are: your specific cognitive resources. A book in Russian does not quickly make sense to me, but to another agent carrying around interpretation procedures for Russian it may.

That points in the direction of thin agents. Within every thick agent we expect to find a thinner agent. On the outside of a thin agent we may find a thicker agent, layers wrapped around the thin agent, strengthening its information procuring abilities. Anyway, the division between what is considered internal and what is external is quite arbitrary from an information point of view; roughly, we expect "internal" resources to be closer than "external" resources (but as we all know this is not always true, it may take longer time to recall a certain name or fact than to look it up in a dictionary or on the web), so that the issue resolves again into information being accessible at varying distances. What we get is basically an onion model. However, in breaking down a thick agent, going inwards, it is not evident that the agent must have a single common center; it might also dissolve into several clusters (as in distributed cognition; Hutchins 1995).

For the purpose of measuring information space, it seems we cannot allow the agent to be *too* thin, either: an agent with *no* cognitive abilities would seem to be a self-contradiction. Perhaps we could define a kind of minimal, rather "stupid" but still meaningful, agent that would serve as a minimal probe of information space.

10 Determining the presence of information and individuating information

To measure the distance to some kind of entities you need to know what the appropriate kind of entity is, and how the presence (or absence) of such an entity can be determined.

Leaving the distance aspect aside for a moment, what criteria should we have for a piece of information being available (closely or distantly)? One possible approach is the pragmatic: information as a key element in performing a certain task, something that (at least potentially) makes an important difference in performance. A problem is how to distinguish information thus defined from other means necessary for performing the task. A related problem is to identify what the added element amounts to, *qua information*.

Another possible approach is *informational satisfaction* or *closure*, which takes more for granted, starting on a less basic level. That is, information as the satisfaction of a request, the satisfaction or allayment of the "irritation of doubt" (Peirce 1877). This approach needs to assume a rather distinct interface between having the information and acting on it. What passes through the information interface? We are accustomed to think of information and knowledge in terms of abstract representations such as propositions. A single proposition ('Scott wrote Waverley') satisfies several requests ('What did Scott write?' 'Who wrote Waverley?' 'Give an example of someone who wrote something!' …). These different requests may have differing satisfaction times (depending on how the proposition is implemented) whereas we expect the distance to a particular piece of information to have a definite value. An alternative (with its own problems, no doubt) might be to base the individuation of information on questions rather than answers: an information entity encompasses whatever satisfies a particular request.

11 Conclusion

I have introduced the concept of *information distance* (and virtual availability) in order to begin the exploration of information space, which I think is the most important step, small and simple that it may seem. Immediately this graded measure of availability raises a number of questions by itself, but the main point of this paper has been to suggest and preliminarily examine a rough idea of what the fundamental structure of information space could be like, and possible methods of continuing the exploration. For that purpose I suggested *availability profile* as a way of defining locations in this space, and tentatively identified it with information state. I have taken the first provisional steps to see if and how this agentcentric view can be reconciled with an infocentric view of information space. Some of the problems and issues confronting the information space pioneer were identified. One is the issue of how to individuate information. At this point it is not clear how crucial different choices are to the coherence and resulting structure of information space. Another issue is that of agent "thickness." Again, it may be that this issue does not have to be finally "resolved" in order to be able to get on with the exploration.

References

Hutchins, E. (1995) *Cognition in the wild*. Cambridge, Massachusetts. MIT Press.

Janlert, L-E. (1985) *Studies in knowledge representation*. UMINF-127.85, ISSN 0348-0542. Umeå. Umeå University.

Moran, T. P. (Ed.) (2001) *Context-Aware Computing*. Special Issue of Human–Computer Interaction, Vol. 16, Nos. 2–4.

Peirce, C. S. (1877) *The Fixation of Belief*. In: Popular Science Monthly, 12 (November 1887), pp 1-15.

Toffler, A. (1970) *Future shock*. New York. Random House.