Letters to the Editor

The information highway

Dear Sirs,

We are all surrounded by information, which is pouring in at an ever increasing rate. There is talk about the information highway in practically every day's newspaper; and the general attitude is that this is good for all of us; it has not yet come completely, but it soon will, and everybody will live happily everafter. Has anybody ever stopped to think whether indeed the information highway is going to be an advantage? Has anybody suspected that it might in fact be a disadvantage? Is "progress" to be defined as whatever comes later, or as improvement? I will discuss the problem from the viewpoint of chemical engineering science.

There were times (being sufficiently old, I can still remember them) when one would actually do something unthinkable of nowadays: without having to look for a specific paper one had heard of by some other means, one would actually go to the library, peruse journals, and actually *read* the papers which looked interesting. One could not, of course, read everything; one had difficulty even knowing which papers had been published in journals which were not available in one's library; but one actually learned something from whatever papers one could read. For the benefit of the younger generation, I will add that one used to actually make notes for oneself (by using such antiquated devices as pencil and paper), and one devised clever ways of filing those hand-written notes. There was a quaint device with cards having holes on their top side such that a long needle through the right hole would pull out all but the relevant cards; subject matters such as chemical reaction engineering have been conceived, developed, and made what they are today by using such incredibly naive methods, and with very little information available. The methods were so naive and cumbersome that one was forced to store away only information that one had learned the hard way was significant.

Then came the Xerox copier, and one gave up the time-consuming exercise of actually reading the interesting papers. It was much easier (and apparently much more efficient) to simply make a copy of the paper and file it. It was so easy that files soon grew to occupy large cabinets, rather than a few small boxes of holed cards, and of course one practically never reads a paper once its xerox copy has been carefully filed. However, at least as long as copying machines were not too good, there were some redeeming features: in order to decide that any given paper was worth photocopying, one had at least to read the abstract; and while making xerox copies one could not avoid glancing at the paper (if for no other purposes than to make sure that the copy had come out OK), and one learned at least a little bit of it. Nowadays we have scanners; the paper gets filed in one's computer, and if perhaps sometimes we read some paper in a file cabinet, let us face it, who ever reads files in one's computer? And one may very well avoid to even glance at the paper while the scanner is scanning it.

One still has to decide to scan the paper, a decision which cannot be taken without a minimum of knowledge of what the paper is about. But that is soon not going to be true: now papers will be on the WWW (they are there already), and it is easy to have an innocent looking program which automatically scans keywords on the WWW and files away everything of potential interest — without one having to even know the existence of something of interest. Everything will be carefully filed away, and we will not have the faintest idea of what is filed in our computer.

The practical result of all this is that we are back to 18th century methods — we learn of what is being published by word of mouth. We go to the library to do whatever we will do with a paper (certainly not actually reading it) which we have heard of by word of mouth. Except for that, the information highway has made it so easy to retrieve information and file it, that we do not ever bother to actually find out what the information is. Much less do we make any use of all the information that is being carefully filed away. If we survive as active scientists, it is because we meet colleagues at Congresses and chat with them, find out that there is something interesting somewhere, take notes — so far still with pencil and paper, but soon with portable electronic notebooks and cellular modems we will unlearn how to read and write, and then we will really be back to word of mouth in the literal sense of the words.

There is something else. We can pull out a book printed in 1680, say, and actually read it. Can anybody read something filed on a $5\frac{1}{4}$ inch diskette nowadays? I mean, easily, as easily as one reads a 1680 book? How many of the readers have a gimmick which reads $5\frac{1}{4}$ diskettes? These things were filed ten years ago — and they are, to all practical purposes, unreadable. Granted, if one thought a given file was important, one would have transferred it, at the time when $3\frac{1}{2}$ diskettes became commonplace and transferring information from one type to the other was easy (somewhere there must still be that gimmick which did the trick, but probably our computer will refuse to communicate with such an antiquated thing). In contrast with this, the book printed in 1680 may well have been regarded as irrelevant for over 300 years, and yet it is still perfectly readable! Of course nowadays $3\frac{1}{2}$ inch diskettes are still readable — but for how long, now that we have CDs? The problem is that information is becoming available at an ever increasing rate, because the means for storing it are improving at an ever increasing rate, so that information stored a few years ago has become in practice unretrievable. On any scientist's desk is sitting at least one computer, not dedicated to calculations but being little more than a sophisticated typewriter and an electronic file cabinet; and that computer probably has something like 1 Gbyte of hard disk memory. When we buy such a computer, do we ever stop to think how much information is 1G byte? How many file cabinets would be needed to store the same information in hard copy form? Should we actually have the

corresponding number of file cabinets we would only need a glance at the sheer space they occupy to readily give up all hope of ever reading even a minuscule fraction of the papers stored in them; but when all that information is on the hard disk of our computer, we somehow think that we could conceivably make some use of it. But do not worry, your hard disk will be unreadable five years from now, and so will the CD, or whatever the information wizards will come out with in the foreseeable future. At the time when we will need to dump all the information on our hard disk to the new system, we will hopefully make some stupid mistake and loose most of it in the process (with a high probability of loosing that minuscule fraction which was in fact useful).

E-mail is another example. It was nice when it started. Then came mailing lists. Of course mailing lists existed even before e-mail; but adding to one's mailing list meant that one had actually to send 500 rather than 200 envelopes around the world; one was somewhat careful about adding to one's mailing list. But adding to one's e-mail mailing list does not create any additional burden; push just one button, and the message goes to everybody on the list, be they 200 or 500. So one ends up in a myriad of mailing lists, and one receives so much e-mail every morning that one does not have the time to read it, or if one does read it one does not have the time to do anything else. Even worse, there came the habit that an e-mail letter could easily include (with a > sign) a copy of the e-mail letter which prompted it — in fact one can set one's e-mail program once and for all to automatically do that. Of course, the letter quoted in my message could easily contain its own sub-letter, but of course that is no problem: the sub-sub-letter will appear under a ≫ sign. We are rapidly getting to the point where the last part of an e-mail message will contain N > signs (with N approaching ∞) — and it will be so long that even scanning it rapidly will become impossible. Again, we started by actually reading e-mail, we moved to only scanning it, and we will soon get to the point where we simply file it and never actually read it — just the same as with papers in technical journals. Of course we will have a way of electronically answering e-mail without reading it, which will simply add one more > sign to the messages which go around the world unread by anybody.

Does anybody remember the nice times when one received mail worth reading? Nowadays only advertising, irrelevant official messages, and memoranda travel by mail; soon even those things will not travel by mail any more. Mail had an advantage: if somebody took the trouble of actually writing down something on a piece of paper, put it into an envelope, and have it mailed, there was a high probability that the message had some importance. Then came the FAX — the letter got there immediately, what a great advantage! One used to read one's faxes carefully — somebody thought this was not only important enough to mail it, it was urgent enough to warrant a FAX. Not nowadays: faxes are sent electronically, one does not need to actually put a piece of paper in a FAX machine, so send them — no matter how irrelevant the message. And send it to everybody on the electronic mailing list! It's so easy! The result, of course, is that one does not read one's faxes any more. There are just too many of them. It is true that a fax still somehow conveys the feeling that the message is somewhat more important than an e-mail one, so that we read faxes a bit more carefully than e-mail — which is perhaps why faxes are going out of use in favor of e-mail.

The amount of available information grows continuously, but the amount of information that any one of us can usefully digest does not grow. The point in time when the bottleneck was the amount of available information has long gone by; nowadays the amount of available information exceeds by far what any one of us can usefully digest. It is, as most often is the case in chemical engineering, a question of comparing time scales. We used to spend time (let us call it t_1) on *retrieving* the information we needed; t_1 has been reduced essentially to zero. Nowadays we spend time t_2 trying to sort out the information we really need from the infinitely large amount of information we do have available — and of course t_2 grows as the amount of available information grows, and it has already become larger than t_1. In former times, one's library had *Chemical Engineering Science*, but it did not have the *Ruritania Chemical Engineering Journal*. So we did not read anything in the latter: if our Librarian had decided it was not worth subscribing to, we assumed of necessity that it contained irrelevant information: crude as such a yardstick is, we used difficulty of retrieval as a yardstick of probable irrelevance. Nowadays retrieval of information is so easy that we have lost that yardstick. The result is that we are submerged by such a large amount of irrelevant information which masquerades as relevant that we just give up the hopeless task of finding something relevant in our computer's files. We file away *Chemical Engineering Science* and *Ruritania Chemical Engineering Journal* papers into the same storage and retrieval system, so we get the advantage of being sure that the occasional relevant bit in the latter is filed away, and we pay the price that the very frequently relevant bit in the former is covered by such a level of background noise as to be lost to all practical purposes.

There is one final point which is worth discussing. In former times, the *producers* of information (say for instance a researcher who wrote a technical paper on his research) could not automatically make it available to the worldwide community of potential users. Papers are refereed before being published; some of the chaff is filtered away. Nowadays journals will end up on the WWW, but (one may hope, but one is not quite sure) this will not kill the refereeing system. But anybody can place anything on the WWW; and be sure that, if the information is irrelevant, it will be placed on the WWW with the best available graphics, and in such way that any scanning program will pick it up. Information that just could not get communicated because somebody decided it was not relevant enough will be communicated (is already being communicated). Instead of being guaranteed to face only information that *somebody* other than the person who produced it has regarded as relevant, we are faced with all the information that the producers consider relevant — and who is the producer of information who thinks his own product is irrelevant?

I have written this letter in a light tone, but I think the problem is a serious one. I can see it clearly with my graduate students, who nowadays are invariably computer wizards and spend at least two hours a day at the e-mail console, because I am scared to tell them to conduct a literature search: if I do, the result is simply a few megabytes of hard disk memory. So I do my literature search on my own, filing on hard copy only what I have actually read, and giving the result to the graduate student. Not very efficient, of course, but still better than the few megabytes of unread material.

I face a dilemma: on the one side, I could place this on the WWW or into all my e-mail lists or both; that would be appropriately ironic. On the other side, I could submit it to *Chemical Engineering Science* and risk that it does not get published; that would be self-consistent. I choose consistency over irony.

GIANNI ASTARITA
*Dipartimento di Ingegneria dei Materiali e della Produzione
Universita' di Napoli "Federico II", Piazzale Tecchio, 80125
Napoli, Italy*