Medical literature as a potential source of new knowledge*

By Don R. Swanson, Ph.D.
Professor, Graduate Library School and Center for Information Studies

The University of Chicago
1100 East 57th Street
Chicago, Illinois 60637

Specialized biomedical literatures have been found that are implicitly linked by arguments that they respectively contain, but which nonetheless do not cite or refer to each other. The combined arguments lead to new inferences and conclusions that cannot be drawn from the separate literatures. One such analysis identified one set of articles showing that dietary fish oils lead to certain blood and vascular changes, and a second set containing evidence that similar changes might benefit patients with Raynaud's syndrome. Yet these two literatures had no articles in common and had never before been cited together; neither literature mentioned the other or suggested that dietary fish oil might benefit Raynaud patients. Two years after publication of that analysis, the first clinical trial demonstrating such a beneficial effect was reported independently by others. A second example of literature synthesis, based on eleven indirect connections, led to an inference that magnesium deficiency might be a causal factor in migraine headache. A third example calls attention to implicit connections between arginine intake and blood levels of somatomedins, a potentially fruitful but neglected area of research with implications for the decline with age of thymic function and protein synthesis.

A model and an online search strategy to aid in identifying other logically related noninteractive literatures is described. Such structures are probably not rare and may provide the foundation for a literature-based approach to scientific discovery.

Each scientific article contributes to a web of logical connections that interlace the literature of science. Some of these connections are made explicit through references from one article to another, citations that reflect, among other things, authors' perceptions of how their own work is related to that of others and how it fits into the scheme of existing knowledge. However, there may exist many implicit logical interarticle connections that are unintended and not marked by citations; such implicit links are the focus of this paper. (The word "logical" here is used informally; a "logical connection" is formed by statements that are related by any process of scientific reasoning or argument.)

* This work is supported by USDE grant RA03980028 from the Office of Educational Research and Improvement—Library Research and Demonstration Program.

Scientific articles can be seen as clustering into more or less independent sets or "literatures." Within each set, common problems are addressed, common arguments are advanced, and articles "interact" by citing one another. Distinct literatures that are essentially unrelated are in general "noninteractive" in that they do not cite or refer directly to each other, have no articles in common, and are not cited together by other articles. On the other hand, if two literatures are linked by arguments that they respectively put forward—that is, are "logically" related or connected—one would expect them to cite each other. If they do not, then the logical connections between them would be of great interest, for such connections may be unintended, unnoticed, and unknown—therefore potential sources of new knowledge.†

† The word "knowledge" is used throughout to refer to scientific or empirical knowledge, and in an objective sense ("recorded

For any single article, little general significance can be attached to the choice of references the authors cite or fail to cite. However, to take a "literature" (a set of dozens or hundreds of articles) as the unit of analysis can reveal larger-scale citation failures and patterns that may be of much greater interest. A recent series of journal articles reports examples that show how, and in what sense, new knowledge might be gained through a synthesis of logically related noninteractive literatures—that is, by assembling pieces already published but perhaps never before put together [2–7]. In each case, logical connections not previously noted in print came to light. The present paper provides an overview and discusses implications of that work, develops a model of the literature structures involved, describes experimental findings that support two of the examples, and introduces a third example. The purpose of this work is to encourage others to set out on similar adventures in literature exploration and synthesis.

If two literatures are linked by arguments that they respectively put forward—that is, are "logically" related or connected—one would expect them to cite each other. If they do not, then the logical connections between them would be of great interest, for such connections may be unintended, unnoticed, and unknown—therefore potential sources of new knowledge.

A MODEL PROBLEM

The idea of logically related literatures and linked arguments can be made more explicit with the help of a model. Suppose that one literature reports that, under certain circumstances, A causes B (e.g., drug A alters blood levels of hormone B). Such a causal statement is denoted by "AB." Assume that a second literature reports a similar causal connection, BC (e.g., hormone B influences the course of disease (C). Presumably, then, anyone aware of the two premises AB and BC would notice that A might influence C (denoted "AC"). This syllogistic construction will be taken informally as a surrogate for more complex modes of scientific argument. (The idea of "logical relevance," illuminated by William Cooper and Patrick Wilson, is of related interest [8–11].) For the purposes of the model, a "literature" that contains an argument such as AB is presumed to include essentially all articles advancing that argument; similarly for BC.

An unknown connection can be discovered by reading both literatures and piecing together their respective arguments, but no one reading one literature would be led by reference citations to suspect the existence of the other.

Assume next that these two literatures are mutually isolated, or noninteractive. That is, they have no articles or authors in common, they do not cite each other, and they are not cited together by other articles. Because there is therefore no printed evidence to the contrary, the intriguing possibility arises that no one person is aware of both premises. AC in that case would be an unintended implicit logical connection that might be known to no one at all. Such an unknown connection can be discovered by reading both literatures and piecing together their respective arguments, but no one reading one literature would be led by reference citations to suspect the existence of the other. This article will show by means of examples that such pairs of logically related but noninteractive literatures do exist and can be discovered.

A syllogism is not the only structure that may harbor implicit connections in the above sense. In principle, any chain of scientific reasoning in which different links appear in distinct, mutually isolated literatures may entail undiscovered, potentially new connections. The chain may include links, such as analogies, that can be neither formalized nor described comprehensively. To bring together such literatures is a goal intended to stimulate the searcher to create and perceive a meaningful mosaic, an act in which invention and discovery are inseparable.

To search for literature structures described by the model, consider, for example, the problem of finding published evidence that leads to an undiscovered cause or cure, A, for a given disease C. Take C as known, A as an unknown agent, and let B represent intermediate links of an initially unidentified physiological chain of events that lead from A to C.

The literature of B must have at least some articles in common with the literature of A and with the literature of C, for there is presumed to be published evidence supporting the premises AB and BC. However, the A and C literatures should have few or no articles in common, because connections are sought that have not previously been formulated or discovered. The problem, then, is to identify two literatures,

knowledge") rather than in the subjective sense of "what someone knows." In a Popperian framework, the abstract world of objective knowledge, though created by man, can contain territory that is subjectively unknown to anyone (past or present) and is open to discovery [1].

corresponding to BC and AB, respectively, that reveal the link B and the agent A.

Typically, the A, B, and C literatures each represent thousands or tens of thousands of articles and so are too large to be read and digested by any one person. BC and AB, however, may each represent only a few dozen articles; to find and identify them is a problem more of exploration than assimilation, for in general they are buried in mountains of other literature.

A trial-and-error search strategy

The structure of the model problem suggests guidelines for a search strategy to aid in identifying logically related noninteractive literatures, a strategy intended to stimulate a creative human process of guessing, testing, and exploring. Such a strategy is based on producing many guesses and then rejecting those that are wrong, an idea not unrelated to how scientists work in general. J. A. Wheeler, quoted by Karl Popper, has said, "Our whole problem is to make the mistakes as fast as possible" [12]. Implications of trial-and-error methods for information retrieval, and of scientific inquiry as a model for online searching, have been explored previously [13–14].

Trial-and-error also describes the evolutionary process of variation and selection. Robert Richards has explained the development of science in terms of a natural selection model in which the usual source of variability lies in the "recombination of ideas" [15]. The process employed here, of excising and recombining ideas from scientific literatures in order to generate new ideas, bears an intriguing resemblance to the Richards model.

Two specific examples of a quest for published evidence of an unknown cause or cure for a given disease are discussed next—the disease in the first case being Raynaud's syndrome and in the second, migraine headache. The complete arguments for each example have been published previously [16–17]. A summary of the findings presented here makes clear the relationship of each example to the paradigmatic ABC syllogism.

The search strategy proposed is neither a recipe nor an algorithm. Success depends entirely on the knowledge and ingenuity of the searcher in forming hypotheses about potential logical connections. Hypothesis formation can be aided by systematic exploratory online searching and reading of appropriate medical literature. Scanning or browsing selected lists of titles that have key words in common is helpful as an initial source of clues to possible causal connections. But however one arrives at an initial conjecture, an "error-elimination" function is necessary; the proposed model suggests a means for rapidly identifying and discarding conjectures that are not of interest because they are probably already well known. Two such procedures—the AC intersection test and a co-citation test—are described in the following example.

EXAMPLE 1: RAYNAUD'S SYNDROME AND DIETARY FISH OIL

Raynaud's syndrome is an episodic shutting off of the blood supply to fingers or toes, usually triggered by cold. Although neither cause nor cure is known, various drug therapies have been used with some degree of success.

Scanning or browsing selected lists of titles that have key words in common is helpful as an initial source of clues to possible causal connections.

To find, within the literature, evidence for an unknown cause or cure for Raynaud's syndrome, the model problem can be used as a guide to an online search strategy. Beginning with the Raynaud literature (C), a search proceeds via some intermediate literature (B) toward an unknown destination A.

One of the conjectures explored was based on discovery of a few reports that some Raynaud patients had abnormally high blood viscosity and red blood cell rigidity. As an example of how online title searching and browsing can be helpful, these reports were readily apparent in a MEDLINE‡ search for all records with the word "Raynaud" in the title and with "BLOOD FACTORS" as a subheading in the descriptor field. Further exploration of the literature about blood viscosity and red cell deformability (*exclusive* of any reference to Raynaud's syndrome) led to noticing a few articles reporting that red cell deformability was increased and blood viscosity reduced by dietary fish oil. Scanning a list of titles that had in common both viscosity-related and deformability-related terms was helpful. Thus, one could infer provisionally from the above two sets of articles that fish oil might benefit Raynaud patients. Once that hypothesis was formulated, other blood factors potentially linking fish oil with the Raynaud syndrome were also found—factors that included vascular reactivity, platelet aggregation, and triglyceride levels. The nature of the exploration problem can perhaps be appreciated by noting that these combined intermediate literatures (B), exclusive of the Raynaud and fish-oil literatures, consist of about 30,000 articles.

Thus, provisional or trial values for A and B were chosen as follows: A denotes dietary fish oil, B de-

‡ MEDLINE is a registered trademark of the National Library of Medicine.

notes alterations in various blood factors, and C denotes the alleviation of Raynaud's syndrome. A small part of the total fish-oil literature contains the argument AB (A causes B), and a small part of the Raynaud literature contains the argument BC (B causes C).

Trial values of A might be chosen initially without knowing whether A has any articles in common with C—that is, whether the set of A-articles intersects the set of C-articles. Thus MEDLINE and other databases were searched to determine the number of records with both fish-oil terms and terms designating Raynaud's syndrome—the AC intersection test. Had there been many such articles, the fish-oil hypothesis would have been considered well known, and therefore dropped as a candidate for an unreported cure. Although there were about 2,000 articles on Raynaud's syndrome and 1,000 on fish oils, it was found that there were none at all on both, a discovery suggesting that fish oil as therapy for Raynaud's syndrome may be a novel hypothesis, and so worth further investigation. In general, the absence of an AC intersection is the key to identifying connections that might be unnoticed. The AC intersection test exemplifies what has been called a "negative search" [18-19]. If the destination of the search were known rather than unknown, a trial-and-error process would be unnecessary and the search logic would then correspond to Soergel's method of subsearches [20].

If A and C have no articles in common (i.e., if their literatures do not intersect), further evidence of mutual isolation is sought by conducting a co-citation analysis [21-22]. If any authors had ever previously examined a possible relationship between fish oil and Raynaud's syndrome, it is plausible to suppose that they would have cited articles from each of the two literatures, that is, co-cited the two literatures. If A is to be retained as a candidate for a novel hypothesis, there should be few or no articles that cite articles about A and articles about C together, for such co-citation would imply that the author of the citing article considered the two literatures together, and so was in a position to notice and explore any logical connections that were thereby brought to light. This argument explains why the definition of noninteractive or mutually isolated literatures includes the absence of co-citation. (To determine all co-citations within and between two literatures, each article is searched, using DIALOG SCISEARCH§, for all subsequent articles that cite it. A process of downloading and sorting then reveals all co-cited groups.)

As a final step, the literatures for BC and AB were examined in depth in order to evaluate the complete biomedical argument. In sum, a review of thirty-four articles on Raynaud's syndrome revealed plausible arguments and evidence that abnormally high platelet aggregability, high vascular reactivity, high blood viscosity, and impaired red cell deformability may characterize at least some categories of Raynaud patients. The most successful drug therapies for Raynaud's syndrome produce favorable changes in one or more of these factors. Analysis of twenty-five articles on fish oil revealed plausible arguments and evidence that dietary fish oils can inhibit platelet activity, lower blood viscosity through increasing red cell deformability or reducing blood levels of triglycerides, increase prostaglandin-I3 production within blood vessels, and inhibit serotonin release. The latter two effects tend to decrease vascular reactivity. All such possible effects of fish oils would be expected to reduce vasospasm, or ameliorate its effects, improve microcirculatory blood flow, and so alleviate Raynaud attacks. Although none of these effects could be considered as conclusively established, the combined evidence seemed strong enough to suggest that a direct test of dietary fish oil in Raynaud patients would be worthwhile.

The two sets of twenty-five and thirty-four articles (based on AB and BC, respectively) are "logically connected," that is, they each contain different premises of an implicit scientific argument that leads to the conclusion AC. The logic of the situation is simple, but the problem is far from trivial because the fragments that reveal the logic are scattered and submerged in a sea of literature. The literatures for AB and BC were also remarkably isolated. They had no articles in common, no authors in common, did not cite or directly refer to each other, and were not co-cited significantly.

In short, by piecing together or assembling arguments in print prior to 1986, a reasonable case could be made for suggesting that dietary fish oil might benefit Raynaud patients. A citation analysis, co-citation analysis, and an online search of various biomedical databases, did not indicate that such a connection had been proposed previously in print, and so a review of the argument and evidence in support of the suggestion was published in 1986 [23-24].

Subsequent corroboration of example 1

In April 1988, two years after example 1 was first published, a randomized placebo-controlled clinical trial in thirty-two patients was reported by researchers at the Albany Medical College, apparently the first report of any test of dietary fish oil in Raynaud patients. In that test, fish oil was shown to delay significantly the onset of experimentally induced Raynaud attacks, enhance blood flow in the fingers, and

§ DIALOG is a registered trademark of Dialog Information Services, Inc. SCISEARCH is a registered trademark of the Institute for Scientific Information.

reduce plasma viscosity. Daily dosage of fish oil contained four grams of eicosapentaenoic acid and 2.6 g of docosahexaenoic acid, and was administered for twelve weeks. The authors concluded that "fish-oil supplements in patients with Raynaud's phenomenon may improve tolerance to cold exposure and delay the onset of vasospasm" [25].

The reported test supports the main point of the argument assembled and published in 1986—that fish oil may benefit Raynaud patients. The most important purpose of example 1, however, is to elucidate the more general form, structure, and significance of logically related noninteractive medical literatures. Such structures can identify potentially fruitful but neglected areas of research, and, in principle, may even anticipate or predict experimental discoveries.

EXAMPLE 2: MIGRAINE AND MAGNESIUM DEFICIT

A second example should further illuminate the logical structure of the first example and support the surmise that there must exist many such pairs of mutually oblivious literatures related by implicit, unnoticed connections.

A search strategy similar to that used in example 1 was applied to the literature of migraine headache (C) and led to the conjecture that magnesium deficiency might be a key causal factor. The search destination, A (magnesium), was unknown at the outset. The destination was reached by several routes and, once reached, led to an unexpectedly large number of additional routes or connections.

An analysis of sixty-five articles on migraine (BC) and sixty-three articles on magnesium (AB) led finally to eleven pairs of implicitly connected arguments, accompanied in each case by substantial, even if not conclusive, evidence [26]. The intermediate literatures (B) are identified by the italicized phrases (Table 1).

Each of the eleven pairs of statements is consistent with, and suggestive of, the hypothesis that magnesium deficiency may be a causal factor in migraine. Yet all of the "a" statements are based exclusively on the migraine literature and the "b" statements on the magnesium literature. Although there were about 4,600 articles on migraine and 38,000 on magnesium, there appeared to be almost no mention of magnesium in the migraine literature or vice versa, nor were the two literatures co-cited significantly. A few exceptions to this mutual isolation were discussed in the full report of this example, including a 1985 paper by Burton Altura that does hypothesize that magnesium deficiency may be a causal factor in migraine, a 1986 report that magnesium sulfate injected in five patients failed to abort migraine attacks, and two older anecdotal reports of successful use of magnesium glutamate injections. Altura alluded to a number of factors included among the eleven connections noted [27]. However, his reference to the migraine literature is sparse, and his paper has not yet been cited by others.

Table 1
Implicitly connected arguments for example 2

Argument 1 (migraine literature)	Argument 2 (magnesium literature)
1. a) *Stress and Type A behavior* are associated with migraine.	1. b *Stress and Type A behavior* lead to body loss of magnesium.
2. a) Excessive *vascular tone and reactivity* may increase susceptibility to migraine.	2. b) Magnesium can reduce *vascular tone and reactivity.*
3. a) *Calcium channel blockers* can prevent migraine attacks.	3. b) Magnesium is a natural *calcium channel blocker.*
4. a) "*Spreading cortical depression*" is thought to be implicated in the early phase of a migraine attack.	4. b) High levels of magnesium in the extracellular cerebral fluid can inhibit *spreading cortical depression* in animals.
5. a) There is evidence for a connection between *epilepsy* and migraine.	5. b) Magnesium deficiency may increase susceptibility to *epilepsy.*
6. a) Migraine patients have abnormally high *platelet aggregability.*	6. b) Magnesium can suppress *platelet aggregation.*
7. a) Platelets of migraine patients are abnormally sensitive to *serotonin release.*	7. b) Magnesium can inhibit *serotonin-induced* contractions of vascular smooth muscle.
8. a) *Substance P* may be a cause of head pain in migraine.	8. b) Magnesium can suppress *Substance P* activity.
9. a) Abnormal *prostaglandin (PG) release* can aggravate vasoactivity in migraine.	9. b) Magnesium increases *prostacyclin (PGI2) formation.*
10. a) Migraine may involve sterile *inflammation* of the cerebral blood vessels.	10. b) Magnesium has anti-*inflammatory* properties.
11. a) *Cerebral hypoxia* may play a key role in migraine.	11. b) Magnesium can protect against *brain damage from hypoxia.*

The main objective of the research reported here is to bring to light unnoticed connections in the medical literature. In pursuing this goal, a second objective has emerged, namely to call attention to hypotheses by others whose work occupies a unique niche in the history of how two separate ideas came together, work that stands almost alone as a bridge between two previously noninteractive literatures, as does the paper by Altura for migraine and magnesium. Moreover, a more thorough literature analysis and additional supporting argument developed in this example led to a number of plausible connections not noted by Altura.

Corroboration of example 2

Four months after the first report of example 2 appeared in print, Kenneth Weaver called attention to an earlier report of an uncontrolled clinical test of

Table 2
Implicitly connected arguments for example 3

Argument 1 (somatomedin literature)	Argument 2 (arginine literature)
1. a) *Growth hormone* stimulates the production of SmC in humans.	1. b) Intake of arginine stimulates the release of *growth hormone.*
2. a) SmC levels in blood are lower in older adults, and lean body mass declines with age. Both growth hormone and SmC are biologically active and can have a wide range of anabolic effects in adults, including the promotion of *protein synthesis, nitrogen retention,* supranormal *growth,* bone turnover, and increase of *lean body mass.*	2. b) Dietary arginine stimulates *protein synthesis* in brain and muscle, improves *growth* and *nitrogen retention* following protein depletion in rodents. In treatment of growth retardation, and emaciation and asthenia in adults, arginine has been reported to promote increased *body weight and muscular strength.*
3. a) SmC promotes *wound healing* after *burns and injuries,* and may enhance *natural killer-cell activity.* Growth hormone has been shown to enhance *thymic structure* and synthesis of *thymic hormones.*	3. b) Arginine intake has been shown to promote *wound healing* after *burns and injuries,* prevent post-injury nitrogen and body-weight loss, and enhance *thymic and lymphocyte response, natural killer-cell activity,* and other aspects of immunological functions.
4. a) SmC levels in both humans and animals drop during fasting and return to normal with refeeding; plasma SmC is a sensitive marker of *nutritional repletion,* especially in cases of *protein–calorie malnutrition.*	4. b) Serum levels of arginine are low in both infant and adult patients with *protein-calorie malnutrition.* Dietary arginine may be of special importance in the *nutritional repletion* of such patients.

supplementary magnesium (200 mg/day) in 3,000 female adult migraine patients [28–30]. An 80% favorable response was claimed, certainly interesting and welcome support for the main point of example 2. Weaver stated, probably correctly, that his was the first and only such clinical test in North America. One hopes, however, that his test will be followed up by a controlled trial. The only report of Weaver's work, except for an abstract, seems to be unfindable through normal bibliographic channels—it has not been cited in SCISEARCH, nor does it appear to be covered by any indexing or abstracting service, either printed or online. Although the Raynaud/fish-oil study (example 1) can be considered a successful prediction of a clinical finding, in the case of example 2 the predicted finding already existed. Perhaps example 2 might be thought of as a successful *retro*diction, having led to a previously existing but undetected, uncited, and virtually unfindable report.§§

EXAMPLE 3: SOMATOMEDIN C AND ARGININE

There is a complex network of implicit connections between dietary amino acids, particularly arginine, and blood levels of somatomedin C (SmC) (also called insulin-like growth factor I) in adult humans, but the two corresponding literatures do not significantly interact. Table 2 summarizes the indirect connections that are reported more fully in a review, published separately, of 127 articles on somatomedins and sixty-five articles on arginine [31].

Each of the "a" statements is based exclusively on the somatomedin literature and each of the "b" statements on the arginine literature (Table 2). The italicized phrases show common elements in the two literatures. Each a–b pair implicitly suggests of SmC, and that such an increase may lead to certain health benefits.

Notwithstanding the foregoing connections, the two literatures on arginine and somatomedins, respectively, are remarkably isolated so far as their citation pattern is concerned. A co-citation analysis of 127 articles on somatomedins and sixty-five articles on arginine was conducted. About 3,000 articles cite one or more of the somatomedin articles, and about 1,000 articles cite one or more of the arginine articles. Only thirty-five articles co-cite at least one somatomedin article and one arginine article. The text of each co-citing article was examined to determine whether and in what sense any relationship between arginine and somatomedin C is discussed. Most of the thirty-five articles are about somatomedins, with only a brief allusion to a well-known stimulation test of pituitary function that uses infused arginine; virtually none of the articles discuss SmC in a context that suggests any direct connection with arginine. Four exceptions to the mutual isolation of these two literatures are discussed in the full report. Judging from the very few citations to these four articles, the idea of investigating a possible connection between arginine and SmC has been largely ignored.

This third example, unlike the first two, does not begin with a disease, but rather with a physiological parameter—blood levels of SmC. There is good reason, however, to suspect a close connection between SmC deficits and a number of disorders. SmC is intimately involved with growth, protein synthesis, and probably with cell maintenance and repair. Moreover, there is evidence that SmC levels, skeletal muscle mass, and protein synthesis all decline with age; such a correlation may be more than coincidental. SmC is a marker of nutritional repletion, and is generally low in cases of malabsorption and malnutrition.

AIDS, especially in more advanced stages, is often accompanied by severe progressive malabsorption and malnutrition. Although there is considerable material on malnutrition and malabsorption in the somatomedin literature and in the AIDS literature, a MEDLINE search (in April 1989) revealed no records common to both somatomedins and AIDS. It is very likely, therefore, that these two literatures are noninteractive and that the question of SMC levels in AIDS patients has not been investigated.

It is of interest to note that statements 2ab, 3ab, and 4ab do not fit the model ABC syllogism (Table 2). Let A represent dietary arginine, B represent somatomedin, and C represent nutritional repletion, wound healing, and other anabolic effects. A new pattern then emerges in which AC and BC are known (that is, arginine has anabolic effects and somatomedins have anabolic effects). The new connection suggested is that possibly B mediates the effect of A on C—that is, perhaps somatomedin mediates certain anabolic effects of arginine (AB). A and B in this case are the noninteractive literatures and AB (A causes B) represents the unknown implicit connection.

In sum, the foregoing arguments suggest that a controlled clinical trial of the possible effect of arginine intake on blood levels of somatomedins may be worthwhile, particularly in the elderly and in patients with emaciating diseases such as AIDS.

A CONNECTION EXPLOSION

The Raynaud/fish-oil, migraine/magnesium, and somatomedin/arginine connections may or may not survive future experimental tests, but whether they stand or fall is less important than where they point. The number of possible pairs of literatures (that might be connected) increases very nearly as the square of the number of literatures; the number of possible connections increases at even a greater rate if triple or higher combinations are taken into account rather than just pairs. From this perspective, the significance of the "information explosion" may lie not in an explosion of quantity per se, but in an incalculably greater combinatorial explosion of unnoticed logical connections. Science responds to growth by increasing specialization, but tends to neglect connections. Inevitably a point will be reached, or may already have been reached, at which connections become of great importance. This emergent property of very large literature bases—the information mosaic of science—presents unique problems and unique opportunities.

FUTURISTIC SYSTEMS: PROBLEMS AND PROSPECTS

The complete published arguments for examples 1, 2, and 3, were assembled from the biomedical literature and did not bring to bear specialized knowledge of medicine or physiology beyond what was explicit in the reviewed publications themselves. Indeed, such a process is intended to cut across specialties, and would not be feasible if expert knowledge of all specialties were required. Although lack of subject expertise may be a limitation in this kind of study, and form and structure of logically connected arguments are in general recognizable by scientists irrespective of their specialty, a point that may have implications for research on futuristic, more fully automated systems. However, the simple structure of the syllogistic model does not in many respects reflect the depth or range of actual problems that would be encountered if one tried to build a database of logical connections.

The significance of the "information explosion" may lie not in an explosion of quantity per se, but in an incalculably greater combinatorial explosion of unnoticed logical connections. Science responds to growth by increasing specialization, but tends to neglect connections.

First, it is clear from the examples that B does not represent a single effect of A, nor does any single effect necessarily cause C. In each example, B stands for a multiplicity of connections, some of which are indirect or linked to others. Such connections are widely scattered in the literature, and very few, if any, are conclusively established, applicable under all conditions, or free of dispute. In none of the examples can one consider a single connection to be as significant as the mosaic formed by many connections. The objective, moreover, is not simply to draw mechanistic logical inferences, but rather to determine whether certain plausible connections or hypotheses appear to be worth testing.

Most articles harbor, either explicitly or implicitly, an enormous number of logical connections. Which connections are relevant and important can be determined only in the light of a specific context, hypothesis, problem, or question; but such contexts and questions are always changing. The degree to which one can hope to encode logical connections in any form suitable for all future uses may therefore be quite limited.

The foregoing problems may be relevant to work on automatic text-processing and knowledge-based systems in artificial intelligence. The approach to literature exploration outlined in this paper may be helpful in defining problems and requirements for futuristic systems, and particularly in illuminating the important role that citation patterns might play in such systems. In this light, the extensive work on co-citation analysis and on the "mapping" of scientific specialties by means of co-citation clusters, appears to hold promise [32–35].

IMPLICATIONS FOR INDEXING

Perhaps a more practical and certain route to futuristic systems lies in the continual incremental improvement of existing systems. A point of departure for improved coding of causal connections may be found, for example, in the present MEDLINE indexing procedures. To a limited extent, certain causal or logical connections are now recognized and recorded by assigning qualifiers or subheadings. There are seventy-six different subheadings that can be attached to specified categories of main headings, which serve to identify the particular aspect of the main heading indexed. About fifteen subheadings express causal relationships, such as "ADVERSE EFFECTS (AE)" attached to the name of a drug, or "CHEMICALLY INDUCED (CI)" attached to a disease heading. The use of certain subheadings in pairs represents a still more explicit approach to the coding of causal connections. AE and CI appearing together in the same record, for example, signifies that a drug that has adverse effects causes a disorder that is chemically induced [36]. Many other such "causal" pairs also are identified in the indexing manual. Subheadings have been of particular importance in the exploratory search process reported here; their further development, and more consistent and more extensive application, would be valuable. They represent important progress toward a general capability that might include linking together different main subject headings (descriptors) when indexing an article and then encoding the specific role of each descriptor in the linked relationship. Links and role indicators, much discussed in the information science literature of thirty years ago, would seem to merit renewed attention in the light of modern online search capabilities.

Links and role indicators, much discussed in the information science literature of thirty years ago, would seem to merit renewed attention in the light of modern online search capabilities.

The idea that human indexing should be improved and extended stands in contrast to suggestions or implications that it should be replaced by automatic indexing [37–41]. The issues are complex and controversial, but it seems clear that automatic recognition of logical connections in text is probably not feasible. One can gain some idea of the difficulty by attempting to identify through an online text search of titles and abstracts those records to which MEDLINE indexers assigned some given pair of "causal" subheadings [42].

OPPORTUNITIES AND IMAGES

Apart from the question of improving existing systems is that of realizing their present potential. More than 1,000 databases and a half-billion online entries to the world of scientific literature are available now and provide opportunity to all scientists for exploration, synthesis, and discovery far beyond what has yet been put to use [43–44]. Until further advances in artificial intelligence are forthcoming, more diligent application of the real kind might be a good hedge against unexpected delays. The reward system and ethos of science, however, recognize only the physical world as a source of new knowledge. The literature tends to be seen as a sort of knowledge necrology, a mechanism of diffusion that supports laboratory-based discovery, but without a life of its own. Science may be better served by a new image of its literature as a vast mosaic of undiscovered connections, a potential source of countless recombinant ideas—a world with its own endless frontier.

ACKNOWLEDGMENT

Richard L. Landau, M.D., professor of Medicine, and Lewis S. Seiden, professor of Physiology and Pharmacology, have provided valuable guidance on the medical aspects of this project. The author thanks Professor Daniel A. Albert, M.D., for calling attention to reference 25, and Professor William S. Cooper, for a suggestion concerning the indexing of logical or causal relationships between pairs of terms. For many helpful comments and suggestions, the author is grateful also to: Professors Mary Biggs, David C. Blair, Stephen P. Harter, William H. Kruskal, Frederick F. Lighthall, Robert J. Richards, Dagobert Soergel, Elaine Svenonius, Mark B. Turner, Patrick Wilson, and Harriet Zuckerman; and to Eugene Garfield, Henry Small, Fred E. Stafford, Patricia K. Swanson, and George E. Vladutz.

This research has depended vitally on rapid and effective access to library materials, all provided by the superb staff and facilities of the John Crerar Library of the University of Chicago.

REFERENCES

1. Popper KR. Objective knowledge. An evolutionary approach. London: Oxford University Press, 1972:Ch. 3.
2. Swanson DR. Undiscovered public knowledge. Libr Q 1986 Apr;56(2):103–18.
3. Swanson DR. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. Perspect Biol Med 1986 Autumn;30(1):7–18.
4. Swanson DR. Two medical literatures that are logically but not bibliographically connected. J Am Soc Inf Sci 1987 Jul;38(4):228–33.

5. SWANSON DR. Online search for logically-related noninteractive medical literatures: a systematic trial-and-error strategy. J Am Soc Inf Sci 1989 Sep;40(5):356–8.
6. SWANSON DR. Migraine and magnesium: eleven neglected connections. Perspect Biol Med 1988 Summer;31(4):526–57.
7. SWANSON DR. A second example of mutually-isolated medical literatures related by implicit, unnoticed connections. J Am Soc Inf Sci 1989 Nov;40(6):432–4.
8. COOPER WS. A definition of relevance for information retrieval. Inf Storage Retr 1971 Aug;7(1):19–37.
9. WILSON P. Some fundamental concepts of information retrieval. Drexel Libr Q 1978 Apr;14(2):10–24.
10. WILSON P. Situational relevance. Inf Storage Retr 1973 Aug;9(8):457–71.
11. SWANSON DR. Subjective versus objective relevance in bibliographic retrieval systems. Libr Q 1986 Oct;56(4):389–98.
12. POPPER KR. Conjectures and refutations: the growth of scientific knowledge. New York: Harper Torchbooks, 1968.
13. SWANSON DR. Information retrieval as a trial-and-error process. Libr Q 1977 Apr;47(2):128–48.
14. HARTER SP. Scientific inquiry, a model for online searching. J Am Soc Inf Sci 1984 Mar;35(2):110–17.
15. RICHARDS RJ. Darwin and the emergence of evolutionary theories of mind and behavior. Chicago: University of Chicago Press, 1987:15–17, 579–93.
16. SWANSON, Fish oil, op. cit.
17. SWANSON, Migraine and magnesium, op. cit.
18. STIELOW F, TIBBO H. The negative search, online reference, and the humanities: a critical essay in library literature. RQ 1988 Spring;27(3):358–65.
19. SWANSON DR. [Letter] RQ 1988 Winter;28(2):287.
20. SOERGEL D. Organizing information: principles of database and retrieval systems. New York: Academic Press, 1985:356–7.
21. SMALL H. Co-citation in the scientific literature: a new measure of the relationship between two documents. J Am Soc Inf Sci 1973 Jul/Aug;24(4):265–9.
22. GARFIELD E. Citation indexes for science. A new dimension in documentation through association of ideas. Science 1955 Jul 15;122(3159):108–11.
23. SWANSON, Undiscovered public, op. cit.
24. SWANSON, Fish oil, op. cit.
25. DIGIACOMO R, KREMER J, SHAH D. Clinical effects of fish-oil fatty acid ingestion in patients with Raynaud's phenomenon. Arthritis Rheum 1988 Apr;31(4)[Suppl]:S34.
26. SWANSON, Migraine and magnesium, op. cit.
27. ALTURA BM. Calcium antagonist properties of magnesium: implications for antimigraine actions. Magnesium 1985 Jul/Aug;4(4):169–75.
28. WEAVER K. [Letter] Perspect Biol Med 1989 Autumn; 33(1):151–2.
29. WEAVER K. Magnesium and migraine: reversible hypomagnesic coagulative angiopathy, hypothesis and preliminary clinical data. (Abstract 47). J Am Coll Nutr 1983 Mar;2(3):287–8.
30. WEAVER K. Magnesium and its role in vascular reactivity and coagulation. Contemp Nutr 1987 Mar;12(3):1–2.
31. SWANSON DR. Somatomedin C and arginine: implicit connections between mutually-isolated literatures. Perspect Biol Med 1990 Winter;33(2):In press.
32. SMALL HG. Co-citation context analysis: the relationship between bibliographic structure and knowledge. In: Proceedings of the Forty-Second Annual Meeting of the American Society for Information Science, Minneapolis, Minnesota, Oct 14–17, 1979. Minneapolis, MN: The Society, 1979:16:270–5.
33. SMALL H. Co-citation context analysis and the structure of paradigms. J Docum 1980 Sep;36(3):183–96.
34. SMALL H. The synthesis of specialty narratives from co-citation clusters. J Am Soc Inf Sci 1986 May;37(3):97–110.
35. SMALL H, GRIFFITH BC. The structure of scientific literatures I: identifying and graphing specialties. Science Studies 1974 Jan;4(1):17–40.
36. CHAREN T. MEDLARS indexing manual, Part II. Bethesda, MD: National Library of Medicine, Bibliographic Services Division, 1983:sec. 19—17.
37. SWANSON DR. Searching natural language text by computer. Science 1960 Oct 21;132(3434):1099–104.
38. SALTON G. Automatic text analysis. Science 1970 Apr 17; 168:335–43.
39. BLAIR BC, MARON ME. An evaluation of retrieval effectiveness for a full-text document-retrieval system. Commun Assoc Comput Mach 1985 Mar;28(3):289–99.
40. SALTON G. Another look at automatic text retrieval systems. Commun Assoc Comput Mach 1986 Jul;29(7):648–56.
41. SWANSON DR. Historical note: information retrieval and the future of an illusion. J Am Soc Inf Sci 1988 Mar;39(2): 92–8.
42. SWANSON DR. Unnoticed connections in the literature of medicine implications for knowledge representation and natural language searching [Abstract]. In: Smith LC, ed. Artificial intelligence: expert systems and other applications. 1988 ASIS Mid-Year Meeting, May 15–18, 1988, Ann Arbor, MI. Washington, DC: The Society, 1988:43.
43. SWANSON DR. Finding scientific literature [Letter]. Am Sci 1988 May/Jun;76(3):231.
44. WILLIAMS M, ed. Computer-readable databases: a directory and data sourcebook. Chicago: American Library Association, 1985.

§§ A recent medical article provides new and direct support for the hypothesis of example 2, that a magnesium deficit may be implicated in migraine. Researchers at the Henry Ford Hospital in Detroit, Michigan, report that brain levels of magnesium were low during migraine attacks, compared to migraine-free controls; reference [6], above, is cited. As in the case of example 1, this new work supports the idea that literature synthesis of the kind described here can predict, or perhaps even stimulate, scientific discovery. Reference: RAMADAN NM, HALVORSON H, VANDE-LINDE A et al. Low brain magnesium in migraine. Headache 1989 Jul;29(7): 416–9.

Received February 1989; accepted March 1989