Cooperative Intelligence Ecosystems: A Proposed Design

CHARLES L. BERNIER

Apartment 4A, 655 Tourmaline Street, San Diego, California 92109

Cooperative intelligence employs effective cooperation among artificial and human intelligences to serve ecosystems comprising authors, editors, employers, professionals, referees, researchers, reviewers, scholars, scientists, searchers, surrogators, translators, and users. Information services provided by cooperative intelligence ecosystems (CIEs) are intended to be comprehensive and continuing. They are believed unavailable as yet from any single source. Intended goals of CIEs include elimination of reading overload, provision of career update, delegation of searches of discovery and updating to experts, enabling decisions to be based on the entire record of the past, and rebirth of personal libraries.

INTRODUCTION

The survival of complex civilizations and organizations likely requires elaborate communication systems. A major source of complexity is the output of research and development that employs the services of increasingly complex communication systems. Better systems should be able to promote prompt sharing of this output in order to maintain the ethics of knowing,[1] ensure the prompt use of new knowledge, and prevent undesired duplication of effort. Unfortunately, reading overload currently is believed to cause decisions, actions, and research to be based on incomplete knowledge of the past. The inadequacy of present-day information systems to overcome reading overlaod has repeatedly been revealed.[2]

Other inadequacies of existing information systems include the following:

(1) Searches of discovery and updating waste precious research time.

(2) Masses of irrelevant material are purchased by subscription, cataloged, indexed, bound, and filed or shelved only to be read little or not at all.

(3) The separation of relevant from irrelevant results of searches takes time. Effective separation may require the consultation of experts, originals, and translations.

(4) Cataloging, indexing, noting, surrogating, and filing relevant material derived from searches in order to facilitate later searches of recall consumes precious time.

(5) Publish-or-perish coercion may interfere with comprehensive, complete work. Better ways for recognizing authors seem available. Authors are rewarded by achievements that can be facilitated through better information systems. The publish-or-perish syndrome is based on the belief that the number of publications is one valid measure of a professional person. The function of publication to advertise and evaluate professionals can likely be improved by the use of ecosystems that provide biographical publications to reveal the number of works published, their titles, the number of citations to these works, and the names of authors of publications that have cited them. Such a biography is superior merely to counting the numbers of papers in evaluating individuals for employment, promotion, and granting of tenure.[3] Such biographies, provided by ecosystems, should help greatly to decrease the cost and waste of publishing little-read papers and of providing reprints that can function mainly as trophies.

GOALS

Intelligence (artificial and human) is revealed and measured by appropriate choice. Appropriate choices enable the achievement of goals. Some very general goals that seem achievable by a cooperative intelligence ecosystem (CIE) are the following:

• Maintain the ethics of knowing and the cognitive authority of administrative authorities, such as boards of directors, CEOs, decisionmakers, and executives, and of cognitive authorities, such as experts, professionals, and specialists, that are required for trusting cooperation.[4]

• Accelerate progress in professions and science and through them the betterment of civilization.

• Enable eventually basing actions, decisions, development, engineering, progress, research, and science on knowledge of the entire record of a highly specific field rather than on only the limited part of the record that, today, can be read, learned, and remembered in the limited time available.

• Provide career update through continuing education that means lifelong learning in order to decrease or eliminate functional obsolescence.

• Delegate nearly all searches of discovery and updating to experts, viz., to indexers, nomenclature authorities, and professional searchers.

• Disseminate selectively surrogates, indexes, originals (works from which surrogates were derived), and translations. Tailormade minihandbooks and minireviews complete with indexes are pictured as being the principal modes of selective dissemination to members of a CIE, each according to his own specifications.

• Create and update profiles of interests of members of CIEs, e.g., authors, searchers, and users.

• Catalog, establish, index, and update complete personal libraries for CIE members in their subject areas of interest. Such libraries would be designed to facilitate searches of recall in which material read is sought again. Rebirth of the personal library is another way for expressing this goal.

• Decrease or eliminate the publication, binding, and filing or shelving of thousands of tons of printed paper that are currently largely unread or rarely read.

• Provide translations of originals and surrogates.

• Furnish authors, on request, with the services of technical writers and editors.

• Notify members of the names of authors who have cited their works.

• Create efficient, cost-effective information systems.

The overall problem is to design CIEs that enable close approximation to, if not the actual accomplishment of, at least, all of these goals.

FEATURES OF COOPERATIVE INTELLIGENCE ECOSYSTEMS

The following proposed design for CIEs is intended to overcome at least the gross inadequacies of current information systems. Of many designs possible for CIEs, the following one has been chosen to attempt the achievement of at least the above goals.

A CIE is an information system proposed to serve a subfield of a discipline. Members of the CIE, the users, are intended to receive only material relevant to their interests so that *all* of the surrogates and all *necessary* originals (from which the surrogates were derived) of the subfield can be read *completely* by members of the CIE in the limited reading time available. A set of related CIEs covering an entire discipline, such as chemistry, can be called a galaxy. Galaxies of CIEs are proposed to try to correct in disciplines all of the inadequacies of current systems in an efficient, acceptable way.

The principal purpose of a single CIE is to update its members on relevant research in its literatures. Additional purposes are proposed later. Possibly literatures other than those of scientific research can also benefit from CIEs.

Cooperating components of a CIE include artificial intelligence (AI), authors, editors, engineers, indexers, referees, researchers, reviewers, scholars, scientists, searchers, students, surrogators, teachers, translators, and users. The components of a CIE are intended to cooperate in order to achieve goals unattainable by individuals working separately.

The scope of a CIE is determined by the subject area that a professional indexer can competently index. This area could turn out to be 10–100 times the area of interest of a single member of the CIE. Competence of indexers depends on a general understanding of the subject area plus an accurate knowledge of the nomenclature and terminology of the field indexed. Indexers need not know details of works indexed. For the initiation of a CIE, a previously defined subject area, such as analytical chemistry or electrochemistry, can be accepted as the initial scope of the CIE. That is, one of the general subject areas into which abstracts have been placed by *Chemical Abstracts* could serve as the starting subject area for a new CIE in the discipline of chemistry. As a selected subject area changes with developments in the literature, the scope of the CIE should change, as should the area of competence of its editors, indexer(s), professional searchers, referees, and reviewers. Thus, the size of a CIE, estimated to be 10–100 times the area of specific interest of any single member, should ultimately be determined by what indexers can index, authors write, readers read, searchers seek, and users use. As the subject area of a CIE changes, its name can also change to reflect progress. With growth, a CIE could subdivide. Although a CIE is intended to supply only material that matches the profiles of its member users, it can supply indexes to all of the material in the CIE—10–100 times larger. This should enable users to broaden their interests and to check the effectiveness of their profiles and to correct them.

Because authors, searchers, and users comprise a set that overlaps, the benefits of CIEs can be enhanced and shared through the cooperation of all members. In our extremely complex civilization, we have discovered that we survive nearly always through trusting cooperation and only very rarely through conflict and confrontation with others. The vital role of trusting cooperation in the survival of individuals and their civilizations, cultures, and organizations has been rediscovered countless times throughout history. This repeated rediscovery may, in part, today be due to the pervasiveness of news that has been defined by media as that which is exceptional. Pervasiveness of the exceptional makes it seem usual. As a result, trusting cooperation may, unfortunately, then come to seem exceptional to the person who views news as normal.

INPUT TO COOPERATIVE INTELLIGENCE ECOSYSTEMS

The two principal inputs planned for CIEs are research results and member profiles. Research results are obtained from manuscripts, published works, cited and citing works, monographs, reports, surrogates, reviews, and other works. For profiles, members disclose to their CIEs, perhaps on forms supplied to them, what they actually do write, seek, read, and use. These forms and their updates are indexed by CIEs to create and update profiles of interests of members. Addenda, corrections, and deletions are expected at any time to keep profiles current. CIEs are planned to be hospitable to all inputs in their subject areas.

Relevant manuscripts can be sought and acquired by a CIE. Such acquisition, by competing with existing journals, could eventually replace some of them. Such projected replacement is not intended to imply that the journal system of scientific communication has not been a major source of progress in the past.

Also, relevant material can be condensed from published journals and reports, analogous to the ways in which they are now abstracted. Original manuscripts received by a CIE should be reviewed or refereed in the usual way. After acceptance, perhaps after revision, they can be edited, and an estimated number of reprints may be printed, perhaps through contract with a journal publisher.

The use of cited and citing works as input into a CIE helps to preserve the effort and skill that authors have invested in locating relevant material.

PROCESSING PLANNED

From profile data input by members, each CIE is planned to create and update their profiles of interests and to notify other relevant CIEs to forward relevant material that turns up in other CIEs of the galaxy and in other galaxies.

Articles, manuscripts, monographs, papers, reports, etc. that are input into a CIE are to be processed by cataloging, surrogating, and indexing the surrogates. Cataloging provides unique labels for each item so that it can be found through the indexes. Surrogating is pictured as creating the highly condensed components[5] of the minihandbooks and minireviews. All input (profile and bibliographic) is indexed to provide index entries that are composed of headings and modifiers (not simply catch or key words) for correlation by AI in providing selective dissemination. The modifiers for research are datum signatures and terse conclusions[5] verbatim. Through correlation of index entries (including citation index entries) derived from surrogates, originals, profiles, and requests, AI continually establishes and reestablishes sets of authors, articles, members, papers, index entries, surrogates, citations, works, readers, reports, searchers, and users all related through common interests and content. The size and specific members of each set are determined by what authors write, searchers seek, and users use rather than by rationally dividing knowledge of the subject area of the CIE into preestablished, labeled subcategories, subclasses, or subsets. Thus, organization within a CIE is established and reestablished after initiation rather than by preestablished subcategories. The number of CIEs in a galaxy is ultimately determined by these same factors. Each galaxy is designed to be hospitable to new CIEs whose size, scope, content, and necessity are largely determined by AI. Factors that can be used by AI programs to determine the size and number of CIEs include relevance of material, reading overload, and the actual interests of members. There is no point in designing for a member of a CIE an output that he has insufficient time to read because of its size. Each CIE is estimated to cover related literatures from 10 to 100 times broader than the interests of any individual member of the

CIE. This comes about because the highly specific field of any member is different from the fields of the other members of the same CIE, and the sum of all these fields is greater than the field of a single member. Members can also join several CIEs in order to increase their subject areas and thus include cross-disciplinary material. It may turn out that the average member joins four CIEs in order to cover his interests, which are almost certain to be cross-disciplinary. Additional scope aids in avoiding inadvertent loss of relevant material that actually could not have been predicted to be relevant. One major way for the prevention of unwanted loss of relevant material is anticipated to be through the use of indexes provided by CIEs that cover all works processed by the CIE.

Reading overload[2] is a major, unresolved problem of existing information systems. CIEs are planned to eliminate reading overload through the combined use of highly condensed surrogates, specialization, and the organization and reorganization of its highly condensed statements.[5] These statements are derived from the *results* of research; they are not derived from subjects researched nor from topics of research. Physical and verbal results can be viewed as the payoff of research, the use of which promotes progress.

Educations deteriorate with age. Nearly all disciplines are so large today that it is now impossible to keep up—even in a recognized subfield of a discipline.[2] This inability to keep up is due to reading overload, which can be revealed by dividing the annual number of relevant works that *should* be read by the number of relevant works that *can* be read in a year. If the *should/can* quotient is greater than one, then reading overload exists. CIEs are intended to relieve reading overload by providing as highly condensed surrogates *only* reading material that is closely relevant to the current interests of members of the CIE, plus a small amount of fringe material that is intended to prevent inadvertent loss. That is, the AI programs can likely be designed to include fringe material by decreasing correlation coefficients required for selection. A rule can be, "When in doubt, include". Totally irrelevant material would, of course, be excluded.

At present, nearly all articles, papers, etc. in most journals and magazines must remain unread by subscribers because there is insufficient time to read all. This is further evidence of reading overload. Also, unfortunately, nearly all articles published in relevant journals are irrelevant to the current interests of their subscribers because the scope of nearly all journals is vastly greater than are each subscriber's current interests. Subscribers to journals have come to expect to read but few of the annual flood of articles published in them. Thus, an enormous amount of unread paper is published, subscribed to, cataloged, indexed, filed, circulated, bound, and shelved. This paper is a heavy load on special libraries and an even heavier load on personal libraries. CIEs are planned to prevent nearly all of this wasteful publication and library processing.

With the assistance of AI, surrogates for current material are organized and mailed as minihandbooks and minireviews to members of the CIE periodically or aperiodically after a number of surrogates have accumulated. Members can be invited to schedule when material is to be mailed to them.

A CIE provides for selective dissemination of surrogates and their indexes in the following way: Members mail to their CIEs essays about their current interests, copies of relevant works, index entries, references to works that they have cited, references citing their works, and references to apparently relevant material that they have actually found to be irrelevant. Indexers of the CIE index all of this material. AI then uses this set of index entries as a key to the profile of interests and exclusions from interests of a member. Some of this material may first have to be acquired by the CIE. Profiles and parts of profiles that are outside the scope of a given CIE are for-

warded by the CIE to other appropriate CIEs in the same galaxy and in other galaxies. These other CIEs then notify the member's CIE of all material relevant to his/her interests as it appears.

Members request the desired arrangment of the material in their minihandbooks and minireviews, specify either frequency or size of shipments that they want, and specify whether or not the indexes mailed to them are to be cumulated and how often. With the minihandbooks and minireviews are also mailed several copies of lists of references to items in these materials. These lists are on bar-coded forms that can be checked or punched out so as to facilitate automated ordering of originals and translations, at cost, from CIEs.

ARTIFICIAL INTELLIGENCE IN PROCESSING

Cooperative intelligence employs intelligences (artificial and human) to do what each does best. At present, the AI component of CIEs seems best for calculating, correlating, data processing, memory, printing, printouts, reliability, selectively disseminating (including selecting, addressing, sorting, and mailing), speed, and helping to establish and maintain personal libraries that provide rapid, convenient access for searches of recall. Human intelligences are presently best for creative writing, editing, indexing, managing, refereeing, researching, reviewing, surrogating (e.g., abstracting, briefing, digesting, extracting, summarizing, translating, and writing highly condensed statements[5]), and using the results of search and research. As AI computer programs are developed that surpass human capabilities, they should take over those functions for which human intelligence is now essential. However, cooperative intelligence for all AI programs is now necessary and may continue to be necessary indefinitely. Cooperative intelligence is not new. No AI system functions completely independently of any human intelligence at some point in its development and use.

AI can contribute, it is believed, to a functioning CIE in several ways. Input to AI includes (1) index entries created, for the present by human indexers, for the highly condensed surrogates that are used to produce the minihandbooks and minireviews, (2) corresponding references to these index entries, surrogates, and originals; (3) useful citations to these originals; (4) names of citing authors; (5) names of members of CIEs; (6) index entries for profiles of members; and (7) computer programs for functions of the AI system. AI uses those correlations that it has determined, through the comparison correlations it makes, to be the most effective. AI correlates the index entries for the profiles of members of the CIE with the index entries for new surrogates that are entered into the system and also with the index entries for citing material. Surrogates for material whose index entries are found by AI to exceed an empirically determined correlation threshold are identified, separated, organized, and reproduced in the form of minihandbooks and minireviews structured according to requests by members. Indexes to all new material in a CIE can be compiled by AI and issued periodically or with each shipment of minihandbooks and minireviews to members. Members can use these indexes to do searches of recall; to make, in the subject area of the CIE, their own searches of discovery and updating; and, at times, to test their profiles and to check the adequacy of processing by their CIEs. Members who suspect that their profiles of interest may be inadequate can be assisted by their CIEs making available to them indexes to the galaxy in which their CIE exists. Such comprehensive indexes can be made available through special libraries or by reproducing relevant index entries under specified headings. Until profiles have been made adequate, members may be well advised to check the completeness of their minihandbooks and minireviews by means of their own searches. For some

members, it may be that their requirements and research are of such a nature that adequate profiles cannot be created, even from a cluster of CIEs. These members can be supplied with indexes to clusters of closely related CIEs, perhaps in the form of laser-read compact discs, so as to give comprehensive access in minimum space. From works identified by members in these indexes, their CIEs can create minihandbooks and minireviews specifically for them. Indexes to clusters and galaxies are also useful within a CIE itself to enable checking of indexing and of profiles for reliability and to facilitate searches of discovery and updating by professional searchers and by other experts of the CIE. Checking and improving indexing is a welcome opportunity provided to indexers of the CIE. Professional searches are those planned to be requested by members in order to make their knowledge and collections complete back into the past.

Members can be notified by AI of the names, addresses, and telephone numbers of other authors and of the complete citations of those authors whose works have cited their works. This service should facilitate prompt communication with others who are actively interested in the same field. If beneficial, this service can foster collaboration and can create cooperative enterprises that employ to the best advantage all varied skills and knowledge of people in the same specific field—of course, observing essential secrecy and security. This service can lead to new organizations and help to prevent the warehousing of brains.

AI then labels these minihandbooks and minireviews, as well as indexes covering them, with unique cataloging symbols and with members' names and addresses and mails them on schedules specified by members to them.

AI is designed readily to accept updates, corrections, and deletions of members' profiles. Such updates, etc. can be supplied to the AI component of a CIE in the form of relevant index entries to essays, reading material, material read, etc. AI also updates shipment schedules sent in by members. All of those items from minihandbooks and minireviews that have actually been *used* by members can provide, perhaps, the best input to increase accuracy of profiles. The kinds and effectiveness of correlations are anticipated to be improved gradually by experience with the CIE and with its AI component. Index entries for relevant citing works can also be correlated by AI and the condensed surrogates for these works also supplied to members if they and experience reveal this to be valuable. The flexibility and uses of AI are almost certainly much more extensive than the above simple suggestions indicate. Human ingenuity working with AI can likely accomplish more than this brief outline hints. After AI has been developed effectively to replace human intelligences for certain processes, then AI can assume an increasing role in a CIE. For example, after AI has been programmed to index, surrogate, and translate as effectively as can human intelligences, then AI can and should take over all of these operations.

AI should also be able to refer index entries, members, and surrogates to other CIEs of a galaxy that seem relevant for these particular materials. Thus, all CIEs of a galaxy are anticipated to cooperate through AI in keeping track of relevant material and to supply it to members wherever it appears in a galaxy. Intergalactic transfer should be planned in order to avoid the loss of any relevant material.

AI should easily keep track of the indexes mailed to a member and to cumulate them periodically on a schedule requested by the member. All of this service should enable members, through the cataloging provided, to create effective, personal libraries of minihandbooks, minireviews, reprints, copies of works, plus their indexes. After reading completely the minihandbooks and minireviews received from his/her CIE, all the member has to do is order, on forms supplied,

essential originals or their translations from the CIE and then file the minihandbooks and minireviews by their unique cataloging symbols. Upon receipt of an order, the AI component of the CIE then reproduces reprints (or other format) for the member, adds cataloging symbols and addresses, and mails them to the member, who then reads them and files them in his/her library.

Although CIEs should be named, the profiles of members require no technical names to label them, only the name and address of the member plus a processing code that indicates the specific CIE, schedule of delivery, number of updating, etc. Materials mailed to members also require no technical names but do need cataloging data, name of the CIE, name and address of the member, and perhaps a number to indicate the issue.

OUTPUT FROM COOPERATIVE INTELLIGENCE ECOSYSTEMS

The two principal outputs from CIEs are minihandbooks of current data and minireviews of current research conclusions. The minihandbooks are made up of datum signatures (labeled numbers of salient, processed data). The minireviews are made up of organized, highly condensed statements nearly always of conclusions (e.g., of terse conclusions[5]) from results to which authors have come after completion of their work. The units from which minireviews are composed are identical with the units from which major scientific reviews have been composed for at least a century. Members can be encouraged to think of each minihandbook and minireview as a product that expert searchers, indexers, and nomenclature authorities have prepared specifically for them. Members are thus relieved of nearly all searches of discovery and updating for which some of them may be ill prepared because of different languages and the lack of access to current material and its indexes and for which all members certainly can ill afford the time. The output received by a member is terse, accurately reveals the *results* of research, is arranged into an order specified by the member, and is mailed according to a schedule the member chooses. These two outputs could, if desired, be provided to members in printed, magnetic disc or tape, or compact disc or video disc format for inclusion in his/her personal library. Material in magnetic format can be rearranged repeatedly by a member through the use of personal and/or organizational computers with the aid of computer programs supplied by the CIE. Rearrangement is useful, for example, when a user becomes an author or changes his approach to research. Material mailed to members can be cumulated periodically or aperiodically by the member or by the CIE upon request. Each item in minihandbooks and minireviews is labeled by the CIE with an unique cataloging symbol that guides members reliably to the originals from which the surrogates were derived. Also, the minihandbooks and minireviews are cataloged as units for filing in the member's personal library. Members file the minihandbooks and minireviews after they have read them completely. These materials are intended to be read completely because everything in them has been selected specifically to be closely related to a member's interests as revealed by his/her profiles and profile updates.

Two more outputs are indexes and cumulative indexes to the complete holdings of a CIE. Such indexes may cover a subject area that is 10–100 times larger than the area covered by the single profile of a member. The extended coverage provided by the indexes helps members to check their profiles for adequacy, to expand their subject areas of interest, and to check the proper functioning of their CIE by, for example, locating relevant work for which they did not receive a surrogate. Members should, of course, expect to be notified by surrogates of relevant works that they, themselves, have cre-

COOPERATIVE INTELLIGENCE ECOSYSTEMS

J. Chem. Inf. Comput. Sci., Vol. 28, No. 2, 1988   45

ated. Indexes can be mailed periodically. Indexes and subsequent cumulative indexes can also be filed by members according to the unique cataloging symbols put on them by the AI of their CIE. Thus, members are intended to be supplied with relevant material that is convenient for building personal libraries, continually updated, and indexed and cataloged—*all without further request.*

Currently, researcher's precious time is spent in searching; in reading and discarding irrelevant material; in waiting for reprints and translations; in noting, surrogating, indexing, cataloging, and filing relevant material read; and in ordering (perhaps as reprints or tear sheets) those originals that must be consulted. Cataloging and indexing of the relevant copies, reprints, and translations acquired are, at present, added chores that waste time. Services of CIEs are designed to facilitate searches of recall in which material read previously is sought.

As other outputs, members can request, on convenient, bar-coded forms, the necessary originals and translations. Nearly always, time available for reading, learning, and remembering will permit the reading of very few of the originals from which the surrogates in the minihandbooks and minireviews were derived. Originals and translations are planned to be supplied to members at cost. Bar-coded forms are used in order to automate filling requests and mailing by AI. Such bar-coded forms of lists of works can be provided to members in replicate along with a shipment of corresponding minihandbooks and minireviews. Originals may be required, at times, to enhance understanding; to verify data and statements; to discover assumptions and hypotheses; to find details of procedures, protocols, techniques, and apparatus; to validate proofs; to discover errors and mistakes; and, rarely, to repeat the work. Because present-day handbooks and reviews are now very seldom used with consultation of the works from which they were derived, it seems likely that requests for originals from a CIE would, in time, become equally seldom. All of the originals for the items in minihandbooks and minireviews, because of numbers, are not mailed to members along with the minihandbooks and minireviews. Members are not expected to have the time to read all of them, even if they could read about 70 languages fluently. The translation of all originals into one language would be incredibly slow and costly. There is expected to be barely enough time to read, in one language, minihandbooks and minireviews along with all *essential* originals and translations.

Reprints and copies of originals and their translations can be sold to members and paid for mainly by their employers in order to prevent the costly, unethical obsolescence of professional employees.

Another service of a CIE is to provide highly specialized technical writers and editors to assist author members upon request. This service should help to improve the accuracy, clarity, and completeness of research papers.

Yet another service of a CIE is to make searches into the past literature of a highly specific field. This service enables members to extend their backgrounds and to give the CIE additional coverage. Such searches are run by experts of the CIE.

The number of items in a minihandbook or minireview mailed to a member on his/her schedule will likely be determined by three factors: (1) the member's profile; (2) the number of research results coming into the CIE in a given period; and (3) the schedule of shipment specified by the member. Members are expected to adjust the number of items that they receive to roughly the number that they can read in the limited time they have available. Members realize that this reading provides them with career update and continuing education. Members may find that, ethically, they tailor their specialties to the amount that they can actually read so as to

avoid assuming or appearing to assume cognitive authority that they do not, in fact, possess.[4] In passing, it is interesting to note that there are no longer persons who can accurately and ethically be labeled simply as biologists, chemists, engineers, scientists, and the like; all have actually become specialists in these disciplines as literatures have increased beyond the capability of an individual to read them.

Because the profiles of all members are expected to differ, except possibly for collaborative work or by accident, every minihandbook and every minireview may be different. These differences arise from the fact that nearly all, if not all, research projects differ, just as do the backgrounds of researchers. Researchers are careful to avoid preempting another's subspeciality and duplicating his work. Preempting is unethical, and duplicating wastes time.

Members can schedule receipt of the output of the CIE either periodically or aperiodically after a specified number of units of datum signatures or terse conclusions has been accumulated for them. Keeping track of numbers for members is a suitable function of AI, as is keeping track of dates of shipment for those members who choose to receive their shipments periodically. A principal advantage of longer periods of issue, and consequently usually of greater numbers of units in each shipment, is that the suggestiveness of related items increases geometrically with their number. Such suggestiveness aids in creating hypotheses and in seeing more of a subfield before action. The effect is similar to that observed in the assembly of a jigsaw puzzle in which the more pieces of the puzzle assembled, the clearer and more complete the overall picture becomes. A principal advantage of shorter periods of shipment is currency.

The products and services of the CIEs should eventually provide career update so that actions, decisions, developments, progress, and research can be based on knowledge of the entire past of a highly specific field rather than on only the small part of the past that can presently be rediscovered, read, learned, and remembered.

AUTHOR COOPERATION

Authors, members or not, may contribute to a relevant CIE their writings in the form of articles, books, manuscripts, monographs, pamphlets, papers, reports, reprints, reviews, and word-processor tapes or discs. Acceptance of new material by the CIE will depend, as at present, on favorable evaluation by editors, referees, and reviewers. Any advantages (if there are any) of competition among publishers can be retained through existing journals.

Another major way in which authors can cooperate is by supplying highly condensed surrogates, e.g., of admonitions, advocacy, conclusions, data, explanations, results, etc.,[5] that are derived from their works. These condensed surrogates authors can provide at the starts of their manuscripts, perhaps in the place of abstracts. Such condensed surrogates are not about the subjects studied, topics reported, or the author's intentions or plans; nor need they be abstracts, briefs, digests, extracts, or summaries because they may be too long, may contain irrelevant material, and may omit valuable, processed, salient data and conclusions. The best condensed surrogates are believed to be derived rather from salient *results* to which authors have come after concluding their research. It has been found possible to write terse conclusions according to simple rules that can be supplied by the editors of the CIEs. Effective rules have been typed on one side of a sheet of typing paper with examples typed on the other side. Multilingual authors can further cooperate by supplying these highly condensed surrogates in one language specified by the editor. Such author translation avoids subsequent translation with its hazards of errors and mistakes common in the translation of materials

as short as are condensed surrogates. It is of advantage to authors to have their works discovered. It also benefits authors to have their works read, used, and referenced. Authors and users can translate and surrogate. Authors can be invited to write in one language the highly condensed surrogates at the beginning of their manuscripts just the same as they now provide abstracts in one language. It is anticipated that surrogators, usually authors and users, working dispersed, outside of the CIE, can likewise provide surrogates in one language for relevant materials that have no surrogates with them and for those materials published in other languages.

A CIE, in a few ways, resembles a small, highly specific section of an abstract journal for a discipline.

ECONOMIC CONSIDERATIONS

Factors favoring initiation and maintenance of CIEs include (1) the cost of not knowing, (2) avoidance of functional obsolescence of professional employees, (3) the anticipated greater effectiveness of CIEs in comparison to current methods, (4) treatment of career update for professionals as analogous to maintaining and depreciating equipment, (5) support for ethics of knowing and thus support for trust required for cooperative efforts, and (6) the hope of solving the problem of reading overload and relieving the frustration associated with it.

Investment in support of CIEs can be counted as a business expense that prevents the functional obsolescence of valuable, professional employees. Justifications for expenditures maintaining and depreciating apparatus, equipment, machinery, plant, and tools can be extended, by analogy, to professional employees. Present-day information systems that help but that have repeatedly been found not to be completely adequate in updating professionals include formal courses, continuing education, meetings, membership in professional and scientific societies, reference works purchased for them, refresher courses, seminars, special libraries, and subscriptions to journals.

Evidence for the value of knowing is impressive. Universal support for educational systems bears witness to the value that our culture has placed on knowing—usually obtained through formal education and training. It has been recognized as being ethical to know.[1] Conversely, professing unacquired knowledge is unethical—as well as dangerous—especially when professed by administrative authority.[4] The relatively large sums invested in education and training seem easily justified to ensure the survival of our extremely complex civilization. Position descriptions usually include educational requirements.

The investment in obtaining a Ph.D. degree in four years of graduate school is, let us assume, $200 000 when tuition, tuition subsidy, materials and supplies, earnings (including investment income from savings) foregone in order to attend classes, and miscellaneous expenses incurred are all included. Permitting such a massive investment to become obsolete can justifiably be regarded as irresponsible. A machine tool, for example, that cost $200 000 could be depreciated and maintained by the annual expenditure of considerably more than 2% of its price—and with easy justification. Two percent of $200 000 is $4000 a year for each professional Ph.D. If there are 25 million professional people in the U.S. with bachelors, masters, and Ph.D. degrees whose educations can benefit from career update at a mean of, say, $2000 each annually, the total amount available for all galaxies of CIEs of the kind described here should amount to $50 billion a year. Galaxies of all CIEs of the kind described here should cost less—say half this amount annually—to operate. If this is so, then investment in CIEs could yield better than 100% return on investment, based on reasonable upkeep, that, in turn, is based on the price of an education for a professional person. This return would be much greater if the cost of not knowing were also included. Ignorant, obsolescent, and undertrained employees can be very costly because of missed opportunities, increased errors and mistakes, and legal suits for malpractice and criminal negligence. This is likely to be true for all professions, fields, and skills. The cost of not knowing is high and is increasing with organizational complexity—especially in our nuclear age.

Because members receive career update, continuing education, and lifelong learning from their CIEs that help to avoid unethical, functional obsolescence, members should expect to pay a reasonable membership fee to their CIEs for these services. The balance, perhaps 80–90%, required to support the CIE could be paid by organizations employing members of CIEs simply because ignorance and not knowing are so costly. Investments in CIEs should be especially attractive.

Because of the present-day impossibility for everybody to know everything, civilizations have become cooperating assemblies of specialists who must trust each other simply because they are physiologically unable to read, learn, and know all that those whom they must trust already know. This trust is violated if a specialist does not observe his responsibility to keep up in his own field.

> "In most technical fields today, the thoughtful scientist views his obligation to remain well informed with a sense of acute personal distress".[6]

Through the use of current information systems, keeping up seems impossible. Those who doubt that this statement applies to them can resolve their doubts by inviting colleagues to write examinations on current research results in their fields.[7] It seems that the myth that reading overload does not exist dies hard. Perhaps the myth exorcises the frustration of not knowing and of failing to keep up and thus violating the trust essential for professionals. Present-day information systems seem not to have completely solved this problem so as to keep specialists ethical and to enhance vital confidence in all of those who enable our civilization to survive.

IMPLEMENTATION

CIEs of a galaxy can be implemented one at a time at greater cost than when starting a galaxy of CIEs. In starting one CIE, an entrepreneur can choose an especially promising subfield of a discipline. Perhaps a new journal is needed in this subfield. Rather than starting a journal, a CIE can be started in its place. Publishers of journals can cooperate by converting to a CIE, by contracting to supply reprints, and by competition benefiting both journal and CIE. Authors, editors, employees, employers, engineers, indexers, publishers, researchers, scientists, searchers, teachers, and uses can be invited to identify others in this subfield. Citation indexes should be most valuable in identifying others. Those people identified as likely being interested in a given subfield can be approached by a well-drafted prospectus to become members of the projected CIE. If there is sufficient interest, the CIE can be started. Lack of interest would indicate that another subfield can be explored. Established professional and scientific societies could start CIEs instead of initiating new journals. Also, newly formed societies can perhaps start CIEs that will serve their members better than will starting a new journal. Organizations, such as corporations, could start one or more CIEs to provide career update for their professional employees and to improve updating at lower cost than with customary means. Abstract journals can likely be converted into a galaxy of CIEs with less effort, expense, and delay than in any other way. It may be possible to convert abstracts into terse statements and datum signatures required for minireviews and minihandbooks, likely with the help of originals, while abstractors are trained to write highly condensed statements[5] instead of abstracts. Perhaps an abstract journal is the ideal

place in which to start a CIE or a galaxy of them because it already has editors, indexers, surrogators (abstractors), and other personnel. It also has subscribers.

The success of one CIE should be used to promote CIEs in other subfields of a galaxy and in other galaxies. Anticipated benefits should attract support for CIEs from editors, employers, indexers, librarians, translators, professional searchers, professional societies, publishers, scientific societies, surrogators, translators, and users. Scientific and professional societies, through financial assistance from employers, could be sources of capital, initiative, management, and supervision of CIEs.

CONCLUDING REMARKS

Implementation, even if experimental, faces formidible obstacles including publishers; the NIH (Not Invented Here) syndrome that seems based on fear of disclosure of inadequate creativity, ingenuity, knowledge, and wisdom; risk timidity; and the comfortable paralysis of custom. However, anticipated benefits may prevail.

It now seems possible to improve existing information systems by creating CIEs and their galaxies—or reasonable facsimilies thereof.

This paper elaborates a prediction of the advent of CIEs.[8]

REFERENCES AND NOTES

(1) Bernier, C. L. "Ethics of Knowing". *J. Am. Soc. Inf. Sci.* **1985**, *36*, 211–212.
(2) Bernier, C. L.; Yerkey, A. N. *Cogent Communication. Overcoming Reading Overload*; Greenwood: Westport, CT, 1979; pp 31–38.
(3) Bernier, C. L.; Gill, W. N.; Hunt, R. G. "Measures of Excellence of Engineering and Science Departments: A Chemical Engineering Example". *Chem. Eng. Educ.* **1975**, *9*, 194–197.
(4) Bernier, C. L. Book review of P. Wilson's *Second Hand Knowledge: An Inquiry into Cognitive Authority*; Greenwood: Westport, CT, 1983. *J. Am. Soc. Inf. Sci.* **1984**, *35*, 255–256.
(5) Bernier, C. L. "Terse Literatures". *Encycl. Libr. Inf. Sci.* **1980**, *30*, 312–330.
(6) Myatt, D. O.; Upham, T. E. "A Quantitative Technique for Designing the Technical Information Center". *J. Chem. Doc.* **1961**, *1*(3), 18–29.
(7) Bernier, C. L. "Measurement of How Well Professional People Keep Up with Their Technical Literatures". *J. Am. Soc. Inf. Sci.* **1971**, *22*, 292–293.
(8) Bernier, C. L. "Development of Indexing and Indexes". *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 164–170.

Supply of Information on Chemical Reactions. An Advanced, Topology-Based Method

R. FUGMANN,* G. PLOSS, and J. H. WINTER

Hoechst A.G., Postfach 800320, 6230 Frankfurt am Main, Federal Republic of Germany

Several pitfalls and problems that the authors have encountered and—in part—overcome are described. The idea of the atom identification (AI) number, which has been in use in IDC (International Documentation in Chemistry) firms since 1960, has proved an efficient device for optimizing the precision and recall ratios for reaction searches and thus endows reaction information systems with high survival power. In particular, it makes it possible to find reactions in which the required (sub)structures of educts and products are separated from one another by unforeseeable intermediates. It also promises to overleap even the document boundaries through which these structures may be separated in the literature.

INTRODUCTION

In a large majority of the queries concerning chemical reactions, the inquirer has a specific substructure in mind, and he is seeking pathways that are described in the literature as leading to this objective from a different substructure, which should be contained in an educt. An efficient documentation system for chemical reactions must render it possible to formulate such queries without having to place restrictions on the type and number of possible intermediate steps. Nor should it be necessary for the inquirer to restrict himself to the retrieval of those publications in which only the *direct* conversion of the educt substructure into the product substructure is described. No inquirer can really predict which ingenious and useful indirect routes, leading from the educt structure in question to the product structure being sought, have already been described in the literature.

It is only with the help of a search process that disregards intermediate steps that all relevant documents can be retrieved with sufficient accuracy from a reaction file in which not only direct conversions from structure I to structure II are contained, as shown in Figure 1, but also indirect conversion processes are contained, which are also depicted in Figure 1.

Frequently the inquirer feels certain that the reaction being sought proceeds only in one step and can have been described in the literature only in this way, as, for example, the conversion of structure I to structure II in Figure 2. If, however, the capability is available of conducting searches that overleap

intermediate steps, it is always surprising how many multistep pathways for the conversion process are encountered in the literature. An example of this is also shown in Figure 2.

An obvious way to conduct searches that leap the intermediate steps would appear to be that of requiring the co-occurrence of the educt and product structures in the same document. However, such generalized search parameters will lead to many responses that do not satisfy the search objective. For example, anyone requiring the mere co-occurrence of structure I from Figure 1 as the educt and structure II as the product will receive the three irrelevant documents with the contents presented in Figure 3.

Intermediates-overleaping searches have been performed at IDC[1] for many years with the help of the GREMAS system. Here in the process of indexing each reaction, certain criteria are used to determine which atoms and bonds should be considered to be involved in the reaction. These reaction sites are then especially characterized, and at the same time it is determined which carbon atoms in the educt and product "correspond" to each other in the sense that they only represent different reaction stages of one and the same atom. This correspondence is indicated by assigning to them the same, arbitrarily chosen "atom identifying (AI) number". A carbon atom retains this number through all the stages through which it passes in the document concerned. In Figure 1 the numbers entered beside the "reacting" carbon atoms are these AI numbers.