PERGAMON

Information Processing and Management 37 (2001) 91–117

INFORMATION PROCESSING & MANAGEMENT

www.elsevier.com/locate/infoproman

Computing a curriculum: descriptor-based domain analysis for educators

Howard D. White

College of Information Science and Technology, Drexel University, Philadelphia, PA 19104, USA

Received 21 July 1999; accepted 10 February 2000

Abstract

College educators need objective ways of assessing coverage, overlaps, and gaps in courses in their curricula, of validating their present offerings, and of monitoring subject-matter trends. This article presents a new methodology for attaining these goals through the use of descriptors from commercial bibliographic databases. A small set of master terms is chosen to model a college, department, or academic degree program, and then large numbers of descriptors that co-occur with the master terms are downloaded in online retrievals. The number of master terms with which descriptors intersect, and the number of documents these intersections produce, yield weights by which the descriptors' relevance to the curriculum can be prioritized. Curricula are thus grounded in the subject indexing of evolving literatures. Suitably arranged, the descriptors form a rich outline of the subject matter, both central and peripheral, that coursework in a field might cover. From this outline, the descriptors with the highest weights are extracted as a "Virtual Curriculum," against which the subject-matter of existing courses can be validated. If individual courses are assigned duplicate high-weight terms, overlaps in course content become visible. High-weight terms that cannot be matched to any existing courses reveal possible curricular gaps. Because online bibliographic databases are dynamic, domain analyses such as this can be repeated periodically to monitor trends and update judgments. A single analyst can carry out all or much of the work; the main costs are for online searching and the analyst's time. The results are comparable to those produced by a national committee of experts. The study reported here used nine master terms to model the curricula for Drexel University's graduate programs in information systems and library and information science. Descriptors from the INSPEC and ERIC databases were processed with Dialog search software (principally the RANK command) and SPSS.

Keywords: Information analysis; Data mining; Curriculum analysis; Curriculum evaluation; Higher education; Bibliographic databases

E-mail address: whitehd@drexel.edu (H.D. White).

0306-4573/00/$ - see front matter
PII: S0306-4573(00)00013-3

1. Introduction

1.1. Literatures as evidence

When a college-level curriculum is being created or revised, educators need an overview of the world of knowledge that the curriculum will model. If an appropriate overview can be generated, so, presumably, can an appropriate curriculum, through faculty selections. The task is easier if faculty can simply recognize topics for inclusion, rather than having to imagine them. This presupposes that the overview of knowledge is explicit — a consultable outline or list. Presented here is a new way of generating that overview, on a customized basis, by computer. Also presented is a way of automatically generating a list of the most salient topics in the overview as a "Virtual Curriculum." This Virtual Curriculum may be used to validate the actual curriculum of a degree program.

The means by which these products are generated is large-scale processing of commercially available bibliographic data. This is feasible where people are skilled in online retrieval (e.g., through Dialog) and basic statistical computing (e.g., with SPSS). The actual retrieval is inexpensive and need be performed only every few years. It can be done by a faculty member or delegated to librarians if that is preferred (cf. Sayles, 1985). Zhu and Porter (1998) discuss this general form of data mining as "knowledge discovery in databases."

The fundamental idea is that contemporary professional literatures, as reflected in bibliographic databases, yield the best — the fullest, most current, most objective, most intelligible — indicators of the subjects that a curriculum might contain. Although human knowledge in many areas of learning is intractably large, and remains so even when reduced to the compass of what has been written down and published, when these literatures are in turn reduced to the compass of citations, abstracts, and indexing terms, we begin to have tractable models of knowledge from which customized outlines can be drawn.

For curriculum planners, bibliographic data are a largely untapped resource. A goal of the present article is to bring their potential to light for wider consideration. The intended reader is anyone wanting to generate useful information on what up-to-date coursework might cover, on interconnections among courses, and on subject matter in need of new or revised syllabi — in short, on curricular coverage, overlaps, and gaps (cf. Septon, 1983; Buckenham et al., 1986; Wright & Larson, 1990; Glass, 1992; Rosen et al., 1992). On matters of determining or validating the content of one's own curriculum in detail, the educational literature is not very helpful; one sifts through handbooks and reviews such as Toombs and Tierney (1991), Jackson (1992), and Gaff and Ratcliff (1996) without finding a usable methodology. There seems to be no systematic way of modeling knowledge domains, other than trusting faculty to read widely and to monitor developments in professional associations and competing programs (cf. Diamond, 1989, pp. 61–64).

The techniques to be shown here are distantly related to earlier work on curricular concept mapping (e.g., English, 1980; Edmondson, 1993). They may be especially useful for curriculum planners in fields whose boundaries are unsettled and whose content is undergoing rapid change. Insofar as new knowledge appears in publications, bibliographic databases can be used to monitor its evolution (Lancaster & Loescher, 1994; Coulter, Monarch & Konda, 1998). Not only are new writings periodically added to them, but their indexing vocabularies are periodically updated to reflect new subjects. Given uncertainty as to where disciplines and industries are going, these updates point the way, not only through terms that are added, deleted, or revised, but also through terms that remain stable. Terminological change and stability are both important. They reflect where a field is currently — a sounder basis for curriculum evaluation than forecasts as to what will be important in five or ten years. Forecasts are often wrong. It is also usually the case that what academics, employers, and students think of as "futuristic" or "forward-looking" or "the next wave" is already fully present in the literature — and consequently in bibliographic databases, if they would turn to them.

Bibliographic databases are a source of evidence that is independent of any particular committee or college. Faculty members responsible for a curriculum presumably want to ground it in something larger than their own opinions. Those opinions may be the best guide as to what *should* be taught, and the only guide, realistically speaking, as to what *will* be taught. But in developing them, faculty usually welcome views of prospective subject matter that do not depend solely on their own school's history, their own mental habits, their own vocabularies, and their own group dynamics. That is why they sometimes seek extramural evidence — for instance, from employers or other schools' catalogs or recent graduates — of what the curriculum should contain (cf. Reitsch & Nelson, 1990; Dey & Mand, 1992; Litecky & Arnett, 1992; Laribee, 1992; Gambill & Jackson, 1992; Richards & Sanford, 1992; Trauth, Farwell & Lee, 1993; Womble, 1993; Spicer, 1994; Lee, Trauth & Farwell, 1995). The richest sources of evidence, I would argue, are the authors, editors, and indexers of a discipline or field. The literature they create is both the most broadly inclusive and the most easily monitored reflection of that larger world of knowledge that curriculum planners want to address. In this view, *the literature warrants the curriculum*. The literature shows what, at a given time, is teachable; the curriculum shows what is taught.

1.2. Academics and descriptors

Despite some foreshadowings (e.g., Dameru, 1990; Allen et al., 1993; Kanter, Miller, Tan & Schwartz, 1994) content analysis of whole literatures has been, until quite recently, too labor-intensive for most educators to contemplate. That day is now past. Existing software permits automation of the most tedious parts of the process, saving the analyst's energies for interpretation. For those who like large returns on small investments, the results may be intellectually appealing. In the present study, the first of its kind, they might even be called striking, since they seem almost more like what industrious, well-read, idiosyncratic persons might supply than the output of a computer.

Granted, it is hardly yet time to add the computer to the curriculum committee. Human judgment remains essential in the formation, sequencing, and articulation of courses (cf. Diamond, 1989). But for curriculum planning and evaluation, it is not too much to claim that online bibliographic databases can be made to behave like helpful intelligencers. They store, and reveal in an ordered way, numerous useful associations of subject matter that would otherwise have to be cudgeled from committee members under deadlines, if they could be gotten at all.

The difficulty in getting any group of academics to agree on the intellectual organization of any large field — on how its topics of study will be named and arranged — is that professors tend to view the field through the lenses of their own formative years and their own present specialties. Given a classificatory task, they typically overelaborate their own areas and relegate everything else to underelaborated residual categories or dismiss it altogether. If they lack an ear for vocabularies not their own, they may have trouble grasping linkages that are obvious to others. At worst, they simply veto each other's categories and fail to create an explicit consensual scheme. As a result, what passes for "the field" remains unarticulated, a tangle of incompatible visions.

The alternative, using literatures as warrants for curricular planning, is done by surveying their descriptors in bibliographic databases — the terms that indexers assign to publications from a thesaurus of authorized terms. The fullest, least biased description of any field is likely to be found in the thesaurus of its abstracting service (there may be more than one). Thesaurus editors are responsible for categorizing literatures on the national or international level. They must select terms that describe the different subjects that authors write about, and, because their schemes will be used everywhere, they must arrange these descriptors evenhandedly, so as to promote consensus. To do so, they must consider the full range of creativity within the field, the documentary evidence in its totality — something most professors lack the time (and, often, inclination) to do. The descriptors they devise do not favor one interest over another, but are arrived at by what librarians call "literary warrant," the promptings of the writings themselves.

A domain analysis based on descriptors can be made to serve many of the same ends as a literature review, but it can be carried out without the very extensive reading and writing that goes into the review, which helps to maintain the analyst's psychological momentum. Thus, during arguments over intellectual content and boundaries, descriptors could serve a wider purpose than they do now. Their potential as global organizers of knowledge is lost to those who associate them only with brief lookups in online literature searching. This is not to claim that they are always up-to-date or that they fit all interests equally; professors can often improve on them within their specialties. But they are almost certainly better globally than anything that might be achieved by local brainstorming.

To be sure, descriptors represent knowledge by noun phrases rather than full statements, and only with the latter can claims of fact or opinion, the real substance of teaching, be made. But noun-phrase descriptors can be linked back to statements in the abstracts or full texts of publications if necessary. Moreover, the language of curriculum design is replete with noun phrases — for example, the names of courses and the topics to be taken up in syllabi — with which the noun phrases from databases can, as a rule, be directly compared.

1.3. Innovation at Drexel

These observations emerge from a study of the curricula at Drexel University's College of Information Science and Technology (IST), which in 1994 received a five-year grant of $1.1 million from the W.K. Kellogg Foundation to innovate in curriculum design and management (Kellogg Project, 1998). To Kellogg, IST proposed an ambitious methodology for a curriculum design that, among other things: (1) derives explicitly from the needs of society; (2) develops and implements new methodologies to support continuous curriculum development, maintenance, and evaluation; (3) blends the requisite disciplines in many combinations to create a range of information and computing professionals; and (4) derives undergraduate, graduate, and continuing education from a common curricular base. Descriptor-based modeling speaks to the goals of the proposal, reintroduced in italics:

- Bibliographic descriptors *derive explicitly from the needs of society*. They operationalize societal requirements as the topics brought to the fore by hundreds of authors. Descriptors are arguably the most comprehensive expression of these requirements. They are also the most neutral, in that they are an external source of information beyond any faculty's control.
- Their use, as evidenced here, *develops and implements a new methodology to support continuous curriculum development, maintenance, and evaluation*. Descriptors help to determine what constitutes a field today, and they evolve with the literature over time, thereby providing a basis for ongoing curricular evolution. They are hard evidence of new developments that curriculum planners should take into account.
- Descriptors can be gathered to any depth or extent required. They can be surveyed or monitored unobtrusively; one need not worry about respondent fatigue. They can be made to reflect more than one discipline, and they can be combined and recombined as necessary to describe various kinds of learning. There is often a choice of bibliographic databases from which to draw them (also, in some fields, a choice between databases of the trade press and of the academic press). In this sense, they *blend the requisite disciplines in many combinations to create a range of information and computing professionals*.
- Descriptors are much the same as the noun phrases of curricular outlines; relatively little new coding is necessary to analyze them for content. Applied consistently, they would support faculty efforts to *derive undergraduate, graduate, and continuing education from a common curricular base*.

In pursuit of these Kellogg project goals, a gathering of descriptors from two bibliographic databases produced results for IST that are novel, relevant, and unexpectedly rich. Perhaps even more surprising, the results agree substantially with those in a national undergraduate curriculum designed by a committee of educators from three major professional associations (*IS '97*, 1997). The real news here may be not so much the agreement — everyone is following the same zeitgeist — as that the substantive judgments of many experts can largely be duplicated at low cost by a single outsider doing computerized literature analysis.

2. Methodology

2.1. Modeling the institution

In the present study, hierarchically ordered descriptors from thesauri are transformed into an Outline of Knowledge and a Virtual Curriculum. However, the descriptors were not obtained "manually," by lookups in the thesaurus. Instead, a small set of master descriptors was chosen to express the essential interests of IST. Then, long lists of descriptors that co-occurred with these master terms in the actual indexing of publications were obtained from bibliographic databases with Dialog software. The online retrieval process, which is both quick and efficient, will be sketched later. For now, it is enough to say that these descriptor lists are *not* what one would get by entering a thesaurus with some terms and copying out everything that is hierarchically or synonymously related to them. The lists instead comprise terms from many different hierarchies, which are linked not because of their a priori meanings, but because they have jointly been used to describe publications. The results are coherent but highly diversified; it is in this respect they seem like what "well-read, idiosyncratic persons" might supply. The following steps for obtaining such results will be elaborated below:

- Know as much as possible about the institutional setting of the curriculum under review.
- Choose databases covering literatures appropriate to the institution.
- Choose a parsimonious set of terms to represent the major themes of the institution.
- Convert the thematic terms into a set of master descriptors.
- Choose a database vendor whose software supports rank-ordering of descriptors.
- Go online to obtain the descriptors associated with those chosen as masters.
- Group the retrieved descriptors into an Outline of Knowledge, whose main headings can serve as an Overview to the nature of the field.
- Extract from the Outline the terms algorithmically identified as most salient to form a Virtual Curriculum.

The entire process will be illustrated with decisions made in studying Drexel's IST. Presumably, analysts can adapt these decisions to other locales.

2.2. Institutional setting

The College of Information Science and Technology (IST), located on the Drexel University campus in Philadelphia, is a somewhat unusual institution, in that it offers a master of science (MS) degree in library and information science, along with curricula in applied organizational computing of a sort more commonly found in business schools or computer science departments. The nucleus of IST is America's third oldest library school, dating from 1892. But, like many of its peers, the college has long sought to expand the employment market for its graduates beyond libraries. In 1984, IST instituted a bachelor of science degree in information systems (BSIS) and in 1993, a master of science degree in information systems (MSIS), both aimed at the nonlibrary employment market. The MS and MSIS programs are those of a professional school; qualified students enter with different undergraduate majors, take a common core of required courses, and then finish their credit requirements with electives. All three degree programs now run concurrently, along with a Ph.D. program dating from 1974. A master's degree program in software engineering, offered jointly with two other Drexel colleges, began in 1997.

The programs at all levels are offered by one undepartmentalized faculty with different backgrounds and specialties. The main split is between those oriented toward library and information science (LIS), with its focus on subject content, and those oriented toward relatively content-neutral areas of computing, like systems analysis, database management, software engineering, or user interface design. In IST parlance, there is an "LIS" side and a "systems" side. The split naturally extends to the students in the different degree programs.

Graduates on the LIS side work mainly in academic, public, special, or school libraries. This side is also called ‘‘information services,’’ since skills found among librarians can also be used in other settings, such as corporate information centers or bibliographic publishing. ‘‘Information systems’’ graduates, who are even more broadly marketable, might go into any sort of organization to shape its computing and networking technology.

2.3. Choice of databases

The goal in modeling a college's or department's world of learning is to capture its major themes in appropriate bibliographic databases. For some proposed curricula, there will be no doubt as to which database to use; only one is appropriate, or one stands out above all others. Sometimes two may seem appropriate (as in the present study), and more might be used. In case of uncertainty as to their identities, it will be necessary to spend some time reading descriptions in the vendor's catalog of databases (Dialog's catalog, for example, indexes databases by broad subject groups, making it easy to select likely candidates).

A second consideration in thematic analysis is the need to be scrupulously fair. It would be impolitic to choose a database that could be seen as favoring one set of faculty interests over another. For IST, this caused the rejection of, e.g., *Library and Information Science Abstracts* on one side and *Computer Database* on the other. Instead, the first choice was INSPEC, the British service whose coverage of various literatures since 1969 accords well with interests on both sides. INSPEC covers journals dealing with information technologies of all kinds and their applications in many fields, including library and information science. For schools whose members publish, a test of accord is whether the database includes their faculty's publications. INSPEC passes this test for both the LIS and the systems faculties at IST. (For data mining of INSPEC in another context, see Zhu & Porter, 1998.)

However, because INSPEC is less well developed on the ‘‘human’’ side, another multidisciplinary database was chosen as a complement: Educational Resources Information Center (ERIC), the American service that has covered both reports and journal articles in education since 1966. ERIC is adequate in identifying technological matters as they apply to education, but it is strong in capturing interests that are social, psychological, institutional, or managerial in nature. It also offers a fuller expression of librarianship, especially in educational contexts, than INSPEC. ERIC picks up publications by IST faculty members (on, e.g., school librarianship or children's literature) that INSPEC misses. Jointly, the two services cover the range of computer and information disciplines.

Both INSPEC and ERIC maintain thesauri in which descriptors are alphabetized and classified. The latter display of hierarchical relationships is often useful to the curriculum analyst. Databases lacking classified vocabularies are likely to require more work on the analyst's part if they are to support curriculum evaluation.

2.4. Choice of major themes

Since the retrieval of associated descriptors depends on it, the analyst's chief discretionary task is to choose the set of terms that will epitomize a particular college, department, or program. This choice of major themes involves judgment; it cannot be reduced to an algorithm. While my choice of thematic terms for IST was intuitive, I sought global terms that implied the subject matter of as many existing IST courses as possible. The selections, shown in Table 1, are for the most part broader in their connotations than actual course titles.

The first two terms simply restate the college's name. The second two are used informally to designate its primary degree programs. The third two are important areas of skill long associated with the college. The fourth two are topics implicated, in one way or another, in almost all its courses. The last is, historically, IST's major employment market and remains the one most readily identifiable. Anyone modeling a college, department, or degree program might be expected to produce a comparable list, of comparable length, in stating themes.

The length of the list of thematic terms will be taken up later. A prior question is: What if someone chose to model IST with terms partially or even wholly different from those below? Differences, after all, are to be expected. If two analysts chose different sets of terms as input, would they not produce different Outlines of Knowledge and hence different Virtual Curricula as output? And does not that compromise the whole procedure by making it altogether arbitrary?

I think not. The reason is the interconnectedness of descriptors across the literatures of a given field — an interconnectedness so great that any reasonable set of master terms has a high likelihood of intersecting with — and retrieving — any other reasonable set. In choosing terms, one is constrained by the controlled vocabulary of the thesaurus and also by conventional ways of describing one's college, department, or program. Therefore, as long as any two analysts both try to capture themes with broad-gauge descriptors they could justify to critical peers, they are both highly likely to tap into the same literatures and to find the same terminological linkages coming to the fore. Indeed, each should retrieve not only the other analyst's input set, but hundreds of further terms in common.

Master terms chosen to characterize an institution should also retrieve *each other*, as a measure of their coherence. That is, they themselves should be interconnected through having been jointly applied to documents by indexers. When the master terms chosen to characterize IST are "ANDed" together online, each indeed retrieves all the others, in both INSPEC and ERIC. The documents retrieved by the ANDings number from 2 to 714.

There is, in short, a great deal of overlap in the descriptors to be retrieved by any reasonable set of master terms. The essential task is to capture this overlap by finding the descriptors that

Table 1
Thematic terms for IST

Information science
Information technology
Information systems
Information services
Information analysis
Information retrieval
Multimedia
User needs
Libraries

co-occur with more than one term in the master set. If one thinks of the master terms as traps set to catch evidence of a field's intellectual structure, then descriptors that are caught by many or all of the traps are the field's core topics. Terms caught less broadly can be taken as defining the field's peripheries. Different sets of master terms will affect chiefly the peripheries; the core concerns should emerge as pretty much the same. In the present study, some descriptors co-occur with all nine master terms in both INSPEC and ERIC. There can be little doubt that the topics they name are central to the information professions that concern IST; the patterns from which meanings emerge are immediately visible. Every field or discipline has similar core topics awaiting discovery.

The question of the length of the master-term list remains. Why *nine* descriptors, as opposed to some other number? Again, the number was reached intuitively. In fact, the actual searches in ERIC and INSPEC made use of slightly more than nine terms, as will be explained. But "about nine" was a parsimonious choice, in that it allowed me to represent IST's major themes without losing the momentum needed for concentrated searching. More important, the set of "about nine" was sufficient for different levels of co-occurrence across retrieved descriptors to be unmistakably observed.

To expand on this last point: the more master terms one uses, the more ties one can break between master terms and the descriptors that co-occur with them. The more ties one can break, the more one can distinguish the top descriptors — those of greatest salience for the curriculum (e.g., DATABASES or STANDARDS). The simplest scoring system might give 0 points when a master term and an ordinary descriptor do not co-occur in the indexing of documents and 1 point when they do. Then if only one master term is used, every descriptor that co-occurs with it will be given a score of 1, and in that sense all are tied. If two master terms are used, some descriptors will co-occur with one, some with the other, and some with both. Those co-occurring with both can be given a score of 2, in contrast to the remainder with a score of 1, and thus some of the original ties on 1 will be broken.

By the time nine master terms have been introduced, the descriptors they retrieve will have cross-scores ranging from 1 to 9, which obviously continues to break ties. At the same time, retrieval on each new master term adds new descriptors to the total set. That is what one wants: a comprehensive pool of descriptors, but also enough distinct cross-scores to distinguish convincingly among them. The descriptors with high cross-scores are now relatively few, and they are clearly being put forward by the literature as salient for the curriculum being modeled. (The procedure used to produce the Virtual Curriculum was a bit more complicated but similar.)

Could a set of master terms retrieve descriptors whose cross-scores were almost all low? That would happen if the master terms reflected fields with largely unconnected literatures — say, physics and neurology. Such a result could mean that the master terms were ill-chosen or that the curriculum was incoherent when compared to the literature (local attempts at interdisciplinarity may have outrun literary warrant).

The goal is simply to produce a good set of terms. Thus, if a descriptor is retrieved at all, it does not finally matter that it would have different cross-scores under different sets of master terms. The main thing is that it be available for consideration in curriculum planning. Moreover, if desired areas of a curriculum are missed by a given retrieval, one can always add a new master term within a database, or add entirely new databases (as I did in adding ERIC to INSPEC), until a satisfactorily broad Outline of Knowledge is obtained. The highest goal should be to assist planners, not to vindicate a particular set of terms. Any less than satisfactory outcome is therefore correctable: inputs can be edited and new searches performed until problems are solved.

In the present study, the nine thematic terms produced 2066 descriptors from two databases for the Outline of Knowledge (1214 from ERIC; 852 from INSPEC) — a large but manageable total set. They also produced more than 300 highly salient descriptors for the Virtual Curriculum. The 300-plus items turned out to be about right for fine-grained comparison with actual IST curricula. Had I continued to add themes beyond nine, many already salient terms of the Virtual Curriculum would have become simply more salient, and the Outline of Knowledge simply bigger. But the two lists are already serviceable; no IST colleague's specialty has been slighted by the retrievals; quite the reverse. Of course, if this kind of domain analysis is to be considered trustworthy, others must find such an outcome to hold with different subject matters and in different settings. But it would seem that these techniques can plausibly be tried on behalf of any postsecondary degree program in any college or department.

2.5. Converting themes to descriptors

We turn now to translating those themes for IST into subject language that will generate an Outline of Knowledge and a Virtual Curriculum. The rule of thumb is to look for good matches to one's chosen thematic terms in a thesaurus of descriptors. Alternatively, one might decide not to preselect thematic terms, but to browse the thesaurus in hopes of finding descriptors of appropriate breadth. In either case, one might need to consider multiple, closely related terms (e.g., User Needs, Information Needs) to capture a theme.

The IST themes, their equivalent descriptors in INSPEC and ERIC, and the years in which the descriptors first appeared in the two databases, are shown in Table 2, along with the sizes of the sets of retrieved documents.

Idiosyncrasies in matching descriptors across databases are justified as follows: (1) INSPEC, the first database searched, used INFORMATION ANALYSIS for various activities that are important in IST (subject analysis, citation analysis, bibliometrics). CONTENT ANALYSIS, connoting computer-assisted identification of thematic materials, was the closest approximation in ERIC. (2) Where INSPEC has the single term LIBRARIES, ERIC has multiple terms; LIBRARY SERVICES was most comparable. (3) "Multimedia," which I thought to be a one-word descriptor in INSPEC, actually invokes the three phrases shown. All were incorporated in my retrieval, and none exactly parallels ERIC's nearest equivalent, MULTIMEDIA INSTRUCTION. (4) My attempt at capturing the idea of people's information needs is something of a patchwork. Not being able to distinguish among ERIC's three terms, I used them all and then backtracked to augment INFORMATION NEEDS in INSPEC with "User Needs," a natural-language term rather than a descriptor. One could undoubtedly be more careful about matching, but mismatches to this extent seemed tolerable in an exploratory study. It is also the case that one cannot expect descriptors in different databases to match exactly; one cannot even expect matched descriptors in different databases to mean exactly the same things (e.g., Information Technology in INSPEC and ERIC). Analysts who require exact matches in input terms might substitute natural-language phrases for descriptors, but there is no guarantee that their meanings will be more stable.

2.6. Choice of database vendor

At this writing, the two online vendors most likely to be used by North Americans for curricular analyses are Questel-Orbit and Dialog. Both offer a variety of databases (with little duplication), and both have appropriate retrieval software, including the essential command for rank-ordering descriptors. The analyst needs to be able to call up a database, form a set of bibliographic records meeting a description, and then rank-order and save the descriptors assigned to the records. Europeans have had access to rank-ordering commands since the mid-1980s through vendors such as ESA/IRS. Questel-Orbit (formerly Orbit) was the first to bring an equivalent to North America; its GET command dates from 1988. Dialog followed with its RANK command in 1992. Dialog and RANK were chosen for the present study. Retrievals on the terms in Table 2 were carried out in July 1995.

2.7. Online retrieval

Dialog indexes more than 450 databases under Dialindex/One Search categories, which create bundles of databases for simultaneous searching. For analyses like those seen here, it is better to search databases separately than simultaneously, because the search sets are often very large, with records and associated descriptors in the thousands. Separate searches may also prevent confusions as to which database is supplying the descriptors.

Master thematic terms will ordinarily be descriptors entered with the suffix/DE. When a resulting set initially comprises several thousand documents, the analyst should consider restricting it to the literature of recent years only. I arbitrarily reduced some sets to documents published since 1990, as noted in Table 2. This reduces the processing time of the RANK

Table 2
Conversion of master terms to descriptors in two databases

Master terms	INSPEC descriptors	Set size	ERIC descriptors	Set size
Information science	Information science (1969)	2581	Information science (1966)[a]	221
Information technology	Information technology (1995)	331	Information technology (1986)[a]	1320
Information systems	Information systems (1993)	1961	Information systems (1966)[a]	460
Information services	Information services (1969)[a]	3642	Information services (1966)[a]	405
Information analysis	Information analysis (1973)	1208	Content analysis (1968)[a]	668
Information retrieval	Information retrieval (1969)[a]	3334	Information retrieval (1966)[a]	825
Multimedia	Multimedia systems (1991)[a] Multimedia communication (1995)[a] Multimedia computing (1995)[a]	4966	Multimedia instruction (1966)[a]	594
User needs	Information needs (1995) User needs (ND)	1681	Information needs (1966)[a] User needs (information) (1966)[a]	1380
Libraries	Libraries (1972)[a]	4003	Library services (1966)[a]	1647

[a] Limited to publications appearing from 1990 through mid-1995.

command (I was not striving for sets of roughly equal size). Once an acceptable set is formed, the next command should be: ? RANK DE CONT, where ? is Dialog's prompt, RANK puts frequencies of co-occurrence in descending order, DE stands for the descriptors (as opposed to other kinds of data) co-occurring with the master term, and CONT specifies a continuous display of the entire set. The default return is limited to the top 50 descriptors, but the analyst may display any number desired, including all of them. Again, however, to keep things manageable, the analyst should consider setting a threshold for the lowest acceptable co-occurrence count. I arbitrarily eliminated descriptors that co-occurred with the thematic terms in fewer than four documents. Consequently, many hundreds of descriptors were dropped, on the ground that they would have greatly extended analytical labors for relatively small gains in information. Even so, 2066 descriptors were retained.

A practiced searcher should be able to do retrievals on ca. nine master terms in one database in less than an hour. The output should of course be saved to disk for processing in the next step.

2.8. Classifying the descriptors

The most labor-intensive part of the analysis comes after the ranked descriptor sets have been saved. That is the manual re-sorting of descriptors into coherent groups so that they form a conspectus of topics that a curriculum might cover. Since the descriptors are downloaded in electronic form, classifying them involves extensive cutting-and-pasting to bring out inherent relationships. For people with a classificatory bent, this is a pleasant task, but (if my case is typical) it may take one classifier several days of on-and-off labor. Even so, no other method is likely to generate a list as comprehensive or detailed with the same expenditure of faculty time. The method is much faster and less arduous than writing a review of the literature, and it produces a large body of descriptors that can be directly mapped onto descriptions of existing courses.

The task of reclassification could in principle be automated, because these descriptors are already classified in hierarchical structures in thesauri (e.g., ERIC's or INSPEC's). A computer program should be able to group them under their original superheadings, as found in an electronic version of their parent thesaurus. To do so would disarm charges of subjectivity and correct errors analysts might make through misunderstanding of the thesaurus sense of a term. But, so far as I know, automatic assignment is not yet possible. Instead, the 2066 descriptors of this study were grouped under headings I created, as in Table 3.

This Overview consists solely of my headings (not the descriptors that give rise to them), and is strictly a device for summarization. In making it, I sought readily justifiable terms, and I was frequently guided by the existing classifications of descriptors in the INSPEC and ERIC thesauri. If not, I let coherent groups of descriptors suggest their own headings and often chose what seemed the most generic descriptors as labels. Other schemes could, of course, be created, but some such organization is needed if the descriptor sets are to be quickly used by colleagues. A few of the groupings contain only one or two descriptors, but numbers much greater than that are commonplace. In fact, it is remarkable how many different subjects are amply represented by descriptors. A scheme as highly differentiated as that in Table 3 is required to do them justice. The nine master terms call forth not merely a discipline or field, but a multidisciplinary world of learning and practice.

2.9. Creating the Virtual Curriculum

The final step in the overall process is to extract from the Outline a Virtual Curriculum —

Table 3
Overview to outline of knowledge for the information and computing field

- General Systems Theory and Automation
- Societal Impact of Information Technologies
  - Automation in Industry and Manufacturing
  - Automation in Libraries, Offices, Homes

- Computers
  - Electronic Circuits
  - Computer Architectures
  - Computer Storage
  - Computer Peripherals
  - Computer Programming
  - Computer Software
  - Neural Networks and Genetic Algorithms
  - Computer Graphics
  - Computer Simulation
  - Distributed Computing
  - Multimedia
  - Hypermedia
  - Virtual Reality

- Computer Applications
  - Computing by Fields
  - Business Applications
  - Health Care Applications
  - Other Applications
  - Microcomputer Applications
  - Data Processing Management
  - Data Processing by Fields

- Information Networks
  - Network Services

- Information Storage Media (Physical)

- Image Processing
- Optical Technology
- Signals
  - Information Theory

- Telecommunications
  - Multiplexing
  - Voice Communication
  - Telephone Systems
  - Radio
  - Mobile Systems

- Television Technology
  - Video Technology

- Speech Technology
- Communication Satellites
- Remote Sensing

- Standards

- Systems Analysis
- Database Design and Management
- Human Computer Interface
- Expert Systems/Decision Support Systems
- Software Engineering

- Mass Media
- Publishing

- Information Science
  - Domain Analysis
  - Content Representation
  - Information Retrieval Systems
  - Information Dissemination Systems
    - Scientific and Technical Information
    - Geographic Information
    - Community Information

- Information Sources
  - Forms of Publication (Textual)
  - Forms of Publication (Graphical)

- Librarianship
  - Institutions
  - Buildings and Interior Design
  - Library Management
  - Cataloging
  - Reference
  - Collection Management
  - Children's and Young Adults Services
  - Interlibrary Relations
  - Bibliographic Instruction
  - Library-related Occupations

- Information Systems and Services (IS&S)
  - IS&S in Industries
  - IS&S in Academic Disciplines
  - IS&S in Education for Academic Disciplines
    - Secondary Level
    - Postsecondary Level
  - IS&S and Educational Levels
  - IS&S and Types of Education

- Users of Information Systems and Services
  - Reading
  - Literacies
  - Specific User Groups
    - Adults
    - Children and Young Adults
    - Students
    - Disabled
    - Disadvantaged

- Management
  - Decision Making
  - Planning
  - Finance and Budgets
  - Quality of Outcomes
  - Personnel
  - Change
  - Public Relations
  - Legal Matters
  - Management Information Systems

- Organizations and Organizational Behavior
- Politics and Government
- Policy Studies
- Public Opinion

- Professionalism
  - Professional Communication
  - Professional Issues
    - Ethical Matters
    - Intellectual Property
    - Multiculturalism
    - Gender
    - Internationalism

- Research and Evaluation
  - Research Methodologies
  - Measurement and Statistics
  - Research in Various Disciplines

- Neighboring Disciplines
  - Mathematics
  - Logic
  - Artificial Intelligence
  - Linguistics
  - Cognitive Psychology

- Educational Matters
  - Educational Institutions
  - Parents and Families
  - Teachers
  - Programs
  - Students
  - Communities
  - Classrooms
  - Schools
  - Colleges
  - Curriculum
  - Training
  - Tests
  - Textbooks
  - Outcomes

- Educational Technology
- Instructional Delivery Modes

those descriptors most highly associated with college or departmental themes. To identify core topics, curriculum planners can use the weights that are automatically attached to descriptors in the course of retrieval. One possible weight is the *number of master terms with which a descriptor co-occurs*. (The more terms, the more central the descriptor.) Another is the *number of documents associated with each co-occurrence*. (The more documents, the more central the descriptor.) The two weights can be combined by summing *documents across co-occurrences*. (The higher the sum, the more central the descriptor.) This last option was used to select topics for the Virtual Curriculum.

For every descriptor a retrieval score was obtained by combining the two weights. These scores were placed in a frequency distribution. It is typical of such distributions to be highly skewed rather than normal, and the one in the present study was no exception. But a suitable transformation (log normalization) produces a distribution much closer to normal, and one can then create standard scores (Z scores) for all the descriptors. Both operations are easily performed in statistical packages such as SPSS. The descriptors most important to IST were obtained simply by asking for all descriptors with Z scores of at least $+1$ — that is, all that were at least one standard deviation above the mean. This reduced the set of descriptors from 2066 to 332. In effect, terms sending out signals of any kind to IST (as long as they made the four-document threshold) were reduced to those sending out *strong* signals.

Table 4 displays the raw data from which these signals emerge. It is a page of a very large SPSS printout in which the nine master terms are cross-tabulated with the descriptors, here from ERIC, that they retrieved. In the cells are the number of documents that each pairing of a master term and a nonmaster term would yield. The extent to which descriptors conform to the nine-theme model of IST can be seen at a glance. Clearly, the terms CURRICULUM DEVELOPMENT and DATABASES, which would produce sizable retrievals across all nine master terms, are central for the college. Note that the row total for DATABASES is 717, giving it a much higher weight even than CURRICULUM DEVELOPMENT, whose total is 211. DATABASES, in short, are here predicted to be a powerful interest across the board in IST, on both the systems and the LIS sides. In contrast, terms like CULTURAL EXCHANGE and CULTURAL IMAGES (top rows) are weak and peripheral.

Table 4
SPSS cross-tabulation of master terms and retrieved ERIC descriptors (edited sample page)

	MASTER TERMS									
ERIC DESCRIPTORS	INFO ANALYSIS	INFO RETRIEVAL	INFO SCIENCE	INFO SERVICES	INFO SYSTEMS	INFO TECHNOLOGY	LIBRARIES	MULTI MEDIA	USER NEEDS	Row Total
CULTURAL EXCHANGE							4			4
CULTURAL IMAGES	9									9
CULTURAL PLURALISM							21		6	27
CURRICULUM	4				6				7	17
CURRICULUM DESIGN	6					9		6		21
CURRICULUM DEVELOPMENT	20	9	14	5	12	58	35	36	22	211
CURRICULUM EVALUATION	11					4				15
CURRICULUM GUIDES	5									5
CURRICULUM RESEARCH	4									4
DATABASES	6	179	10	49	70	117	98	31	157	717
DATABASE DESIGN		34			10				16	60
DATABASE MANAGEMENT		34			11	14	7		6	72
DATABASE PRODUCERS		7								7
DATA ANALYSIS	6	5			5		6		21	43

In a part of the SPSS printout not shown, the data roundly support IST's long-planned master's degree program in software engineering. The descriptor SOFTWARE ENGINEERING has substantial co-occurrences in INSPEC with eight of the nine master terms, especially INFORMATION SYSTEMS.

Table 4 also reveals the attraction of the present methodology for those who must explain it to others. It is apparent that columns hold vocabularies associated with the master terms singly, while the rows show descriptors interacting with the master terms jointly. It is also apparent that the cell counts and the row totals can be used to further distinguish between descriptors. At a presentation, audiences can be shown the patterns for descriptors of interest to particular individuals. They can also be shown the patterns for descriptors that faculty groups have in common (perhaps without being unaware of it). For example, the descriptor STANDARDS (i.e., conventions to promote uniform procedures), which happens to be used by both INSPEC and ERIC, produces sizable retrievals across all nine master terms in both databases. This suggests for IST that standards are an important subject — a transcurricular subject — for both LIS and systems faculty.

3. Results

3.1. The outline and its overview

3.1.1. The overview

The world of the Overview is centered on IST's professional curricula. Table 3 is arranged so that the middle column contains the central subjects of the College — information systems and services. The two flanking columns reveal the larger relations of study. The left column expresses the harder side of information technology — computing and telecommunications — and its headings for the most part come from literatures covered by INSPEC. The right column expresses relatively nontechnological subjects, and a much greater proportion of its headings reflect literatures covered by ERIC. Literatures from both INSPEC and ERIC contribute large numbers of headings to the middle column.

The three columns together indicate the range of topics of interest to the College and to other colleges at Drexel (e.g., the left column suggests Drexel's programs in electrical engineering, computer engineering, and computer science; the right column, its programs in technical and scientific communication, business management, and education). IST faculty had already perceived these intramural linkages, of course, but this analysis of bibliographic data reinforces their intuitions with documentary evidence. Table 3 could be used to show administrators that the existing allocation of subject matter at Drexel — and at many other universities — makes sense. It could also be used as an objective representation of academic turf to be divided.

The Overview reveals that the retrieved descriptors go far beyond being close synonyms of the input terms. Two major kinds of nonsynonymous linkages appear: conversion of the input terms to more specific hyponyms (e.g., "Multimedia" to "Virtual Reality") and, strikingly, associations (such as those to "Multiculturalism" and "Gender") that are not part of the predictable lexical field of any input term.

The bundles of descriptors assigned to the headings in the Overview take up more than 10 densely printed manuscript pages, too many to include here. One of these pages, however, is reproduced as Table 5. It elaborates the headings "Systems Analysis" through "Information Sources" from the middle column, so that the reader may see how the Overview derives from and is warranted by the full Outline. I shall return to the Overview shortly.

3.1.2. The outline

The full Outline includes many more topics than any IST curriculum will ever address. It may be thought of as a 2066-item menu of potential subject matter from which faculty can choose the topics they will actually cover. (The Virtual Curriculum, extracted from the Outline, permits a more focused test of the subject coverage of existing courses.) Table 5 provides the reader with illustrative descriptors and weights to be checked for face validity in rendering subject matter for a college such as IST. Each descriptor is preceded by a count of the master terms with which it co-occurs in INSPEC (I) or ERIC (E), nine being the top score. (Recall that co-occurring terms must retrieve at least four documents to be included here.) The descriptors are ranked by co-occurrence frequency across the nine terms. When both INSPEC and ERIC produce identical descriptors in a retrieval, these will sometimes be found together in a ranking; if not, the higher is followed by a "1" and the lower by a "2" to indicate they are a separated pair. The number of master terms with which descriptors co-occur is one measure of how important they are for the IST curriculum, present or potential. The ranked counts prioritize descriptors, so that faculty can see which topics are being weighted by various literatures as most relevant to the curriculum. In each grouping, therefore, look particularly at the top-ranked terms, especially those with co-occurrence frequencies of, say, 6 or higher. The latter are important for curriculum planning because the higher they are, the more they pervade the thematic definition of the college or department.

The importance of the co-occurrence counts as term weights cannot be overemphasized. When weights of all terms in the Outline are compared to a threshold value, many of the hardware-related descriptors in the left column of Table 3 *drop out* as suitable topics for IST curricula. So do many descriptors in the right column (e.g., most of those under "Educational Matters"); they make the first cut, but not the second. And while the center column has the greatest proportion of high-weight terms, some of its topics drop out also. The weights thus serve as a corrective to the subjectivity in the arrangement of Table 3 (my sense of subject similarities, my labeling, my layout in columns of equal length, and so on). Even if the Overview is eliminated as a summarization device, the term weights remain to show topics of greater or lesser import to the curriculum. The effect is to highlight indicators of the disciplinarity of the College in the context of its interdisciplinary relations.

An analysis of more than 2000 descriptors provides a great deal of detail about potential subject matter. Table 5 reveals typical levels. It will be seen that the descriptors under such headings as "Human–Computer Interface," "Software Engineering," and "Content Representation" begin to resemble outlines for the individual lectures in one or more courses. Rather than vague generalities, there is considerable concrete suggestion for teachers here, and of course the descriptors can be plugged into online searches to retrieve the specific writings on which they rest. They may also evoke topical sections in new or standard textbooks.

Table 5
Outline of knowledge for the information and computing field (sample page)

Systems Analysis		
9	I	SYSTEMS ANALYSIS 1
9	E	NEEDS ASSESSMENT
8	E	COMPUTER SYSTEM DESIGN
6	E	DESIGN REQUIREMENTS
5	E	SYSTEMS DEVELOPMENT
5	I	FORMAL SPECIFICATION
5	I	TECHNICAL SUPPORT SERVICES
5	I	MAINTENANCE ENGINEERING
3	E	SYSTEMS ANALYSIS 2
3	I	SYSTEMS RE-ENGINEERING
3	I	OPTIMISATION
2	I	SECURITY
1	I	SYSTEMS ENGINEERING
1	E	TASK ANALYSIS
Database Design and Management		
9	E	DATABASES
8	I	DATABASE MANAGEMENT SYSTEMS
8	I	DATA STRUCTURES
8	I	SECURITY OF DATA
7	I	DATA HANDLING
7	I	QUERY LANGUAGES
7	I	RELATIONAL DATABASES
6	I	OBJECT-ORIENTED DATABASES
6	I	QUERY PROCESSING
5	E	DATABASE MANAGEMENT SYSTEMS
5	I	DATA ACQUISITION
5	I	DATA ANALYSIS
5	I	DATA PRIVACY
5	E	DATA PROCESSING
5	I	DATA VISUALISATION
5	I	FILE ORGANISATION
4	I	VERY LARGE DATABASES
4	I	DATABASE THEORY
4	I	DATA COMPRESSION
4	I	DATA INTEGRITY
4	I	ABSTRACT DATA TYPES
4	I	DEDUCTIVE DATABASES
3	I	SQL
3	E	DATABASE DESIGN
3	I	ENTITY-RELATIONSHIP MODELLING
3	I	TEMPORAL DATABASES
2	I	TREE DATA STRUCTURES
2	I	QUERY FORMULATION
1	E	DATABASE PRODUCERS
1	I	DATA CONVERSION
1	I	DATA ENCAPSULATION
1	I	TABLE LOOKUP
Human Computer Interface		
9	I	USER INTERFACES
8	I	INTERACTIVE SYSTEMS
6	I	GRAPHICAL USER INTERFACES
6	I	REAL-TIME SYSTEMS
5	I	NATURAL LANGUAGE INTERFACES
5	I	ONLINE FRONT-ENDS
4	I	USER INTERFACE MANAGEMENT SYSTEMS
4	E	MENU DRIVEN SOFTWARE
2	E	SCREEN DESIGN (COMPUTERS)
2	I	ONLINE OPERATION
1	I	COMPUTER INTERFACES
1	I	TOUCH SENSITIVE SCREENS
1	I	LIQUID CRYSTAL DISPLAYS
1	I	INTERACTIVE DEVICES
1	I	INTERACTIVE PROGRAMMING
1	I	INTERACTIVE TERMINALS
1	E	KEYBOARDING (DATA ENTRY)
Expert Systems/Decision Support Systems		
8	I	EXPERT SYSTEMS 1
8	I	DECISION SUPPORT SYSTEMS
8	I	KNOWLEDGE BASED SYSTEMS
7	I	KNOWLEDGE ACQUISITION
7	I	KNOWLEDGE ENGINEERING
7	I	KNOWLEDGE REPRESENTATION
5	E	EXPERT SYSTEMS 2
3	I	MEDICAL EXPERT SYSTEMS
1	I	DIAGNOSTIC EXPERT SYSTEMS
Software Engineering		
8	I	SOFTWARE ENGINEERING
6	E	COMPUTER SOFTWARE DEVELOPMENT
6	I	SOFTWARE TOOLS
5	I	SOFTWARE PROTOTYPING
5	I	SOFTWARE REUSABILITY
4	I	SOFTWARE PERFORMANCE EVALUATION
3	I	SOFTWARE QUALITY
3	I	COMPUTER AIDED SOFTWARE ENGINEERING
2	I	RELIABILITY
2	I	FAULT TOLERANT COMPUTING
2	I	SYSTEM RECOVERY
2	I	SOFTWARE AGENTS
2	I	SOFTWARE MAINTENANCE
2	I	SOFTWARE PORTABILITY
2	I	SOFTWARE RELIABILITY
2	I	DEVELOPMENT SYSTEMS
1	I	PROJECT SUPPORT ENVIRONMENTS
1	I	SOFTWARE COST ESTIMATION
1	I	SOFTWARE DEVELOPMENT MANAGEMENT
1	I	SOFTWARE LIBRARIES
1	I	SOFTWARE MANAGEMENT
1	I	SOFTWARE METRICS
1	I	FORMAL VERIFICATION
Publishing		
8	I	PUBLISHING
8	E	PUBLISHING INDUSTRY
8	E	ELECTRONIC PUBLISHING
8	I	ELECTRONIC PUBLISHING
5	E	AUTHORS
4	I	OPTICAL PUBLISHING
3	E	DESKTOP PUBLISHING
3	I	DESKTOP PUBLISHING
2	E	FACULTY PUBLISHING
1	E	EDITORS
Information Science		
9	E	INFORMATION SCIENCE
7	I	INFORMATION SCIENCE
7	E	DOCUMENTATION
7	I	DOCUMENT HANDLING
4	E	INFORMATION PROCESSING
		Domain Analysis
7	I	INFORMATION ANALYSIS
6	E	CONTENT ANALYSIS
5	E	BIBLIOMETRICS
4	E	CITATION ANALYSIS
		Content Representation
5	I	ABSTRACTING
4	E	ABSTRACTS
1	E	ABSTRACTING
9	E	LITERATURE REVIEWS
8	I	REVIEWS
3	I	SOFTWARE REVIEWS
2	E	COMPUTER SOFTWARE REVIEWS
2	E	BOOK REVIEWS
1	E	STATE OF THE ART REVIEWS
7	E	CLASSIFICATION
7	I	CLASSIFICATION
1	E	CLUSTER GROUPING
7	E	ANNOTATED BIBLIOGRAPHIES
6	E	BIBLIOGRAPHIC DATABASES
5	E	BIBLIOGRAPHIC RECORDS
5	E	BIBLIOGRAPHIES
4	E	CITATIONS (REFERENCES)
8	E	INDEXING
8	I	INDEXING
8	I	THESAURI 1
8	I	VOCABULARY 1
8	E	DEFINITIONS
6	E	SUBJECT INDEX TERMS
6	I	GLOSSARIES 1
5	I	NOMENCLATURE
4	E	THESAURI 2
4	E	INDEXES
4	E	CHECK LISTS
3	E	VOCABULARY 2
1	E	GLOSSARIES 2
1	E	LEXICOGRAPHY
1	E	CITATION INDEXES
1	E	AUTOMATIC INDEXING
Information Retrieval Systems		
9	E	INFORMATION RETRIEVAL 1
9	E	ONLINE SEARCHING
9	E	ONLINE SYSTEMS
8	I	INFORMATION RETRIEVAL 2
8	E	SEARCH STRATEGIES
8	I	BIBLIOGRAPHIC SYSTEMS
8	I	INFORMATION RETRIEVAL SYSTEMS
7	E	RELEVANCE (INFORMATION RETRIEVAL)
7	I	INFORMATION RETRIEVAL SYSTEM EVALUATION
5	E	ONLINE VENDORS
5	E	BIBLIOGRAPHIC UTILITIES
2	I	RELEVANCE FEEDBACK
Information Dissemination Systems		
9	E	INFORMATION DISSEMINATION
8	I	INFORMATION DISSEMINATION
8	I	TECHNOLOGY TRANSFER
7	E	DELIVERY SYSTEMS
6	E	INFORMATION TRANSFER
5	I	DOCUMENT DELIVERY
4	E	ADOPTION (IDEAS)
3	E	SELECTIVE DISSEMINATION OF INFORMATION
2	E	TECHNOLOGY TRANSFER
1	E	DIFFUSION (COMMUNICATION)
1	E	LINKING AGENTS
		Scientific and Technical Information
8	E	SCIENTIFIC AND TECHNICAL INFORMATION
6	I	MEDICAL INFORMATION SYSTEMS
1	I	SCIENTIFIC INFORMATION SYSTEMS
1	I	ENGINEERING INFORMATION SYSTEMS
		Geographic Information
7	I	GEOGRAPHIC INFORMATION SYSTEMS
6	I	CARTOGRAPHY 1
2	I	SPATIAL DATA STRUCTURES
1	E	CARTOGRAPHY 2
		Community Information
5	I	PUBLIC INFORMATION SYSTEMS
3	E	COMMUNITY INFORMATION SERVICES
2	E	OCCUPATIONAL INFORMATION
1	E	SOCIAL SERVICES
Information Sources		
8	E	INFORMATION SOURCES
7	E	REFERENCE MATERIALS
7	I	FACTOGRAPHIC DATABASES
6	I	VISUAL DATABASES
6	E	FULL TEXT DATABASES

Table 5 points up the relevance of this sort of domain analysis to tests of curriculum coverage. Glancing over the descriptors, faculty may recognize specific topics they think important — e.g., SECURITY OF DATA under "Database Design and Management" — and are reminded to ask what existing courses cover them and to what depth. Possibly two or more courses cover them; possibly none. In this way, faculty committees might discover unsuspected overlaps or gaps in the curriculum.

The SECURITY OF DATA example is one of many that illustrate the claim, "Bibliographic descriptors derive explicitly from the needs of society." Perusal of Table 5 (and, even more, the entire Outline) should reveal that the underlying literatures are oriented toward the world of practical affairs. While these descriptors would doubtlessly retrieve some writings that are abstrusely academic, they would also retrieve many that are quite down-to-earth. In any case, the descriptors themselves often specify, in plain English, the requirements of present-day business and technology. Terms naming the kinds of knowledge and skills that employers ask for in want-ads are indeed present in the full Outline: for example, C PROGRAMMING, OBJECT-ORIENTED PROGRAMMING, LOCAL AREA NETWORKS, TECHNICAL PRESENTATION, DP MANAGEMENT, FAIR USE (COPYRIGHTS), DISTANCE EDUCATION, CLIENT-SERVER SYSTEMS, and INTERNET.

There is another validity test for such an Outline, and that is its congruence with model curricula at the national level (assuming they are available). As it happens, a model curriculum for four-year undergraduate programs in information systems was released not long after the descriptors in the Outline were obtained. *IS '97* (1997) has a detailed synopsis of topics that can be directly matched against the descriptors as retrieved. Table 6 presents one of the several spreadsheet pages that are needed to effect the match. It will be seen that the two delineations of subject matter are quite close — in many cases virtually identical. I have simply filled in enough blanks to show the general viability of the process; persons with more specialized subject expertise could probably make rough to excellent matches over the entire matrix. I would stress here, as earlier, that the work of a committee of experts can be largely anticipated by one analyst with the requisite retrieval skills and appropriate databases.

Even so, any individual professor who tried to create a model of potential subject matter such as the Overview would almost surely be challenged as to why certain items were included. For example, Professor X might think that topics under "Logic" or "Mathematics" are important to the IST curriculum in information systems, or that topics under "Research Methodologies" are important to the library degree. But if she urged them for the curriculum on her own, her colleagues might see her as privileging her own interests, and that objection would recur even if she used a published thesaurus of descriptors to make her choices. Such perceptions of bias lead academics to endless squabbles. The attraction of working with literature-based descriptors is that the *literatures* say what topics are important in relation to IST's central concerns, and the literatures are the work of hundreds of independent observers. In this case, the literatures, represented by linked descriptors, say that all three of these topical areas are important to IST, and this endorsement is based on "the field's consensus" rather than on the agenda of any local faculty member. There are times when a consensual view of this sort might be valued for its impartiality.

Some headings in Table 3 simply remind one of the ubiquity of information systems and

Table 6
IS'97 outline (first page) matched with descriptors from INSPEC or ERIC

	IS '97 Terms	*INSPEC and ERIC Terms*
1.1.	**Computer Architectures**	COMPUTER ARCHITECTURE. RECONFIGURABLE ARCHITECTURES
1.1.1.	Fundamental Data Representation	
1.1.2.	Physical Representation of Digitized Information	
1.1.3.	CPU Architectures	CONTENT-ADDRESSABLE STORAGE. REDUCED INSTRUCTION SET COMPUTING, etc.
1.1.4.	Computer System Components	PERIPHERAL INTERFACES, etc.
1.1.5.	Multiprocessor Architectures	PIPELINE PROCESSING. PARALLEL PROCESSING. DISTRIBUTED PROCESSING
1.1.6.	Digital Logic & Systems	LOGIC PROGRAMMING
1.2.	**Algorithms & Data Structures**	ALGORITHMS. DATA STRUCTURES
1.2.1.	Formal Problems & Problem Solving	PROBLEM SOLVING
1.2.2.	Basic Data Structures	DATA STRUCTURES
1.2.3.	Complex Data Structures	DATA STRUCTURES
1.2.4.	Abstract Data Types	ABSTRACT DATA TYPES
1.2.5.	File Structures	FILE ORGANIZATION
1.2.6.	Sorting and Searching Data Structures & Algorithms	SORTING. SEARCH PROBLEMS
1.2.7.	Algorithm Efficiency, Complexity & Metrics	ALGORITHMS
1.2.8.	Recursive Algorithms	ALGORITHMS
1.2.9.	Neural Networks & Genetic Algorithms	NEURAL NETS. GENETIC ALGORITHMS
1.2.10.	Advanced Considerations	COMPUTATIONAL COMPLEXITY, etc.
1.3.	**Programming Languages**	PROGRAMMING
1.3.1.	Fundamental Programming Language Structures	PROGRAMMING
1.3.2.	Machine & Assembly Level Languages	
1.3.3.	Procedural Languages	C LANGUAGE
1.3.4.	Non-Procedural Languages	
1.3.5.	Fourth-Generation Languages	HIGH LEVEL LANGUAGES
1.3.6.	Object Oriented Extensions to Languages	OBJECT-ORIENTED LANGUAGES
1.4.	**Operating Systems**	OPERATING SYSTEMS (COMPUTERS)
1.4.1.	Architecture, Goals & Structure of an Operating System	OPERATING SYSTEMS (COMPUTERS)
1.4.2.	Interaction of Operating System & Hardware Architecture	OPERATING SYSTEMS (COMPUTERS)
1.4.3.	Process Management	PROCESSOR SCHEDULING. SYNCHRONISATION. CONCURRENCY CONTROL
1.4.4.	Memory Management	MEMORY ARCHITECTURE
1.4.5.	Resource Allocation & Scheduling	RESOURCE ALLOCATION. SCHEDULING
1.4.6.	Secondary Storage Management	STORAGE MANAGEMENT
1.4.7.	File & Directory Systems	FILE ORGANIZATION
1.4.8.	Protection & Security	SECURITY OF DATA
1.4.9.	Distributed Operating Systems	DISTRIBUTED PROCESSING. DISTRIBUTED CONTROL
1.4.10.	OS Support for Human Interaction	USER INTERFACE MANAGEMENT SYSTEMS
1.4.11.	OS Interoperability & Compatibility	
1.4.12.	Operating System Utilities, Tools, Commands & Shell Programming	OPERATING SYSTEM KERNELS
1.4.13.	System Administration & Management	COMPUTER MAINTENANCE
1.5.	**Telecommunications**	TELECOMMUNICATIONS
1.5.1.	International Telecommunication Standards, Models & Trends	TELECOMMUNICATION STANDARDS. ISO STANDARDS
1.5.2.	Data Transmission	DATA COMMUNICATION
1.5.3.	Line Configuration	ERROR CONTROL. MULTIPLEXING
1.5.4.	Local Area Networks	LOCAL AREA NETWORKS
1.5.5.	Wide Area Networks	WIDE AREA NETWORKS
1.5.6.	Network Architectures & Protocols	COMPUTER NETWORKS. PROTOCOLS
1.5.7.	Internetworking	INTERNETWORKING
1.5.8.	Network Configuration, Performance Analysis & Monitoring	NETWORK TOPOLOGY. PERFORMANCE EVALUATION
1.5.9.	Network Security	SECURITY. AUTHORISATION
1.5.10.	High-speed Networks	B-ISDN. FDDI
1.5.11.	Emerging Networks	ISDN. SATELLITE COMMUNICATION, etc.
1.5.12.	Telecommunications Applications	TELECOMMUNICATIONS SERVICES
1.6.	**Database**	DATABASE THEORY. DATABASES

services in contemporary society. Global descriptors addressing this fact (e.g., SOCIAL ASPECTS OF AUTOMATION) are assembled under the subheads of "Societal Impact of Information Technologies" at top left. Hundreds of other descriptors name specific walks of life affected. For example, the headings under "Computer Applications" in the left column assemble descriptors such as CHEMISTRY COMPUTING, COMPUTERIZED NAVIGATION, PERSONAL COMPUTING, and MARKETING DATA PROCESSING. The headings under "Information Systems and Services" in the middle column link IS&S to about a hundred descriptors naming broad areas of study (HUMANITIES), disciplines (PHYSICS), and subdisciplines (LITERARY CRITICISM), along with preparatory endeavors such as MATHEMATICS INSTRUCTION, ART EDUCATION, HOME SCHOOLING, and CAREER EDUCATION.

The difference between the disciplines assembled under IS&S and those labeled "Neighboring DISCIPLINES" is that descriptors like PHYSICS turn up only once or twice without further elaboration, whereas LINGUISTICS, typifying the neighbors, is elaborated with about 30 more specific terms (e.g., GRAMMARS, SEMANTICS, PHRASE STRUCTURE, NOUNS). The literature is saying that Information Science is more closely related to Linguistics than to Physics, which would surely agree with most people's intuitions in the matter.

As noted above, this methodology produces some striking examples of associations between nonsynonymous descriptors. Who would have suspected that any of the master terms would call up terms related to multiculturalism or gender? Everyone knows that these are salient concerns among present-day educators, but it is intriguing to see them evoked by input terms as neutral and distant as those in Table 1. The full sets of retrievals have been placed in Table 7 to show that they deserve labeled places in the Overview. They are almost all from ERIC rather than INSPEC, but a fair number co-occur with more than one input term. Again, if particular faculty members tried to introduce such concerns into a formal statement of curriculum, other colleagues might see them as advancing a particular ideology. Here the

Table 7
Multiculturism and gender from the outline

Multiculturalism		Gender
6 E Cultural differences	1 E Ethnic stereotypes	5 E Sex differences
5 E Minority groups	1 E Stereotypes	4 E Females
4 E Cultural awareness	1 E Racial bias	2 E Sexuality
4 E Ethnic groups	1 E Racial discrimination	1 E Feminism
4 E American Indians	1 E Racial relations	1 E Sex bias
4 E Hispanic Americans	1 E Asian Americans	1 E Sex discrimination
3 E Multicultural education	1 E Spanish speaking	1 E Sex fairness
2 E Cultural pluralism	1 E Limited English speaking	1 E Sex role
2 E Blacks	1 E Indigenous populations	1 E Sex stereotypes
1 E Cultural images	1 E Popular culture	1 E Males
1 E African culture	1 E Folk culture	1 E Women's education
1 E Black stereotypes	1 E Social bias	1 I Gender issues
1 E Ethnicity	1 E Immigrants	1 E Spouses
1 E Ethnic bias	1 E Civil rights	1 E Marriage

literature is advancing it. The literature evidently reflects contemporary academic life and is politicized in the same way. But for all this to be present in the linkages of dull bibliographic descriptors seems almost to be a manifestation of artificial intelligence.

The AI effect comes not from innovative computer programming but from exploiting massive stores of content analysis by human indexers. The indexers themselves are very likely unaware of the connections they have made; they did not set out to create them deliberately, but simply responded to the literature. But the outcome is that of a successful learning process. The Overview can be read as if it indexed conversations among a very large, diverse curriculum committee over a long period of time. It resembles a huge mnemonic device, designed to promote recognition rather than recall of topics. While I have interpreted it in light of my own college, it could be interpreted to fit numerous other North American schools similar to IST. The broader point is that the methodology seems worth trying by *any* faculty wanting a synopsis of potential matter for interrelated coursework.

3.2. The Virtual Curriculum illustrated

I have called the collection of 332 strong-signal terms a Virtual Curriculum. In Table 8 I have mapped them onto a sample of IST courses (as of mid-1999) at both the undergraduate and graduate levels. Undergraduate and graduate courses that are similar in content are brought together in the listing. The figures to the right of each descriptor are, first, the total number of master descriptors in both databases, from 1 to 18, with which it co-occurs; second, a computed Z score of $+1$ or higher for each term; and, third, the total number of documents in both databases that a descriptor, in combination with the master terms, would retrieve.

In this preliminary mapping, I tried to give each course a bundle of *unique* terms as often as possible (some duplication will be seen, but not much). Most IST courses can be captured by one or more strong-signal terms, and many can be captured even better if additional terms are drawn from the full Outline of Knowledge. I take this to mean that, in general, the literature validates IST's undergraduate and graduate curricula. There are, to be sure, different levels of validation. For example, Database Management I and II are supported by numerous strong-signal terms, whereas some older, established LIS courses, such as Resources in the Social Sciences or Legal Bibliography, are not, and two courses lack strong-signal terms altogether. However, if one looks to the full Outline to find suitable indexing for them, it is almost always there to be found. For example, the full Outline supplies READING INTERESTS, ADULT READING PROGRAMS, RECREATIONAL READING, NOVELS, FICTION, and NONFICTION for the course Reading Interests and Contemporary Literature, aimed at students planning to work in public libraries.

My indexing in any case is intended merely to show faculty how they might build up a subject index for their courses. The work thus far provides them with a vocabulary for doing so that is linked to real thesauri and real literatures. With such an index, one can see how well IST covers the range of subject matter present in evolving literatures, look for overlapping subject matter in courses, and test for gaps in curricula.

Overlap can be good if it implies reinforcement; bad, if it implies redundancy. If faculty index their courses, it is easy to retrieve courses that have the same descriptors. This is a development envisioned under what is called the Drexel Curriculum Workbench, one

Table 8
Virtual curriculum terms mapped onto IST courses (sample page)

Undergraduate courses are labeled INSYS; graduate courses, INFO. Similar undergraduate and graduate courses appear together. Virtual Curriculum terms follow course names. The three columns at right are, respectively, the number of master terms in INSPEC and ERIC with which a descriptor co-occurs (maximum 18), the descriptor's Z score (the sort key), and the number of documents retrieved by the master term and descriptor jointly. In general, the higher these numbers, the more important the topic for the Virtual Curriculum.

INSYS 101 Introduction to Information Systems I			
INFORMATION SYSTEMS	17	3.47	2898
INFORMATION TECHNOLOGY	13	3.32	2346
INFORMATION DISSEMINATION	17	2.68	948
INFORMATION NETWORKS	16	2.60	851
INFORMATION STORAGE	13	1.96	345
MAN-MACHINE SYSTEMS	11	1.56	196
DELIVERY SYSTEMS	7	1.25	126
INFO 502 Information Management Tools			
INSYS 102 Introduction to Information Systems II			
MICROCOMPUTER APPLICATIONS	8	2.51	744
MICROCOMPUTERS	13	2.20	480
WORD PROCESSING	12	1.61	209
COMPUTER LITERACY	11	1.15	110
OPERATING SYSTEMS (COMPUTERS)	6	1.09	101
INFO 503 Introduction to Information Systems Analysis I			
INFO 620 Information Systems Analysis			
INSYS 200 Systems Analysis I			
INSYS 355 Systems Analysis II			
INFORMATION SYSTEMS	17	3.47	2898
INFORMATION TECHNOLOGY	13	3.32	2346
SYSTEMS ANALYSIS	12	2.21	493
COMPUTER SOFTWARE	11	2.15	453
COMPUTER SYSTEM DESIGN	8	1.96	345
SOFTWARE TOOLS	6	1.83	288
NEEDS ASSESSMENT	9	1.60	208
FORMAL SPECIFICATION	5	1.41	159
OBJECT-ORIENTED METHODS	4	1.23	123
INFO 510, 511 Information Resources and Services I and II			
INSYS 105 Information Resources and Their Use			
INFORMATION SERVICES	16	4.01	6208
INFORMATION SCIENCE	15	3.50	3040
SEARCH STRATEGIES	8	2.02	377
USE STUDIES	8	1.74	252
INFORMATION SOURCES	8	1.72	247
LITERATURE REVIEWS	17	1.68	391
FACTOGRAPHIC DATABASES	7	1.48	175
LIBRARY RESEARCH	7	1.44	165
BIBLIOGRAPHIC DATABASES	6	1.43	162
ANNOTATED BIBLIOGRAPHIES	7	1.23	123
PUBLISHING	8	1.16	111
REFERENCE MATERIALS	7	1.07	98
INFO 515 Action Research			
TABLES (DATA)	9	2.10	417
QUESTIONNAIRES	9	1.99	359
STATISTICAL ANALYSIS	15	1.61	211
INTERVIEWS	9	1.51	183
SURVEYS	7	1.47	173
DATA COLLECTION	6	1.20	117
INSYS 215 Social Aspects of Information Systems			
INFO 520 Professional and Social Aspects of Information Services			
TECHNOLOGICAL ADVANCEMENT	9	2.46	700
SOCIAL ASPECTS OF AUTOMATION	8	1.93	331
PROFESSIONAL ASPECTS	8	2.03	382
FUTURES (OF SOCIETY)	9	2.59	843
ECONOMIC AND SOCIOLOGICAL EFFECTS	7	1.49	178
ACCESS TO INFORMATION	9	2.70	979
INFO 601 Computer Programming for Information Processing			
OBJECT-ORIENTED PROGRAMMING	6	1.72	247
INFO 602 Text Processing by Computer			
DOCUMENT HANDLING	7	1.46	169
INFO 605 Database Management I			
INFO 606 Database Management II			
INFO 607 Applied Information and Database Technology			
INSYS 210 Database Management Systems			
DATABASE MANAGEMENT SYSTEMS	13	2.53	770
DATABASES	9	2.48	717
RELATIONAL DATABASES	7	1.91	319
VISUAL DATABASES	6	1.83	287
OBJECT-ORIENTED DATABASES	6	1.70	237
DATA STRUCTURES	8	1.66	226
QUERY PROCESSING	6	1.50	179
QUERY LANGUAGES	7	1.46	171
DATABASE THEORY	4	1.25	127
INFO 608 Human-Computer Interaction			
INSYS 110 Human-Computer Interaction			
USER INTERFACES	9	2.85	1205
INTERACTIVE SYSTEMS	8	2.14	446
INTERACTIVE VIDEO	8	2.07	400
GRAPHICAL USER INTERFACES	6	2.01	372
ONLINE FRONT-ENDS	5	1.65	221
REAL-TIME SYSTEMS	6	1.60	208
INFO 612 Knowledge Base Systems			
INSYS 450 Expert Consultant Systems			
EXPERT SYSTEMS	13	2.20	485
KNOWLEDGE BASED SYSTEMS	8	2.07	401
KNOWLEDGE REPRESENTATION	7	1.47	173
DECISION SUPPORT SYSTEMS	8	1.46	170
DEDUCTIVE DATABASES	4	1.19	116
INFERENCE MECHANISMS	5	1.05	95
INFO 614 Distributed Computing and Networking			
INSYS 350 Distributed Computing and Networking			
COMPUTER NETWORKS	17	3.04	1576
LOCAL AREA NETWORKS	13	2.25	519
PROTOCOLS	6	2.07	400
WIDE AREA NETWORKS	6	1.96	344
DISTRIBUTED PROCESSING	7	1.88	309
ISDN	5	1.82	283
OPEN SYSTEMS	7	1.80	275
BROADBAND NETWORKS	4	1.74	251
INTERNETWORKING	6	1.71	242
DISTRIBUTED DATABASES	6	1.64	220
ELECTRONIC DATA INTERCHANGE	7	1.43	163
DATA COMMUNICATION SYSTEMS	7	1.31	137
B-ISDN	1	1.30	136
PACKET SWITCHING	1	1.14	108
COMPUTER COMMUNICATIONS SOFTWARE	4	1.13	107
CLIENT-SERVER SYSTEMS	6	1.03	93
INFO 616 Computer-Supported Cooperative Work			
INSYS 405 Computer-Supported Cooperative Work			
ELECTRONIC MAIL	15	2.28	543
TELECONFERENCING	13	2.17	464
GROUPWARE	7	2.05	389
OFFICE AUTOMATION	10	2.04	386
WORKSTATIONS	5	1.38	152
FACSIMILE	5	1.05	95
INFO 617 Introduction to Systems Dynamics			
TECHNOLOGICAL FORECASTING	8	1.93	328
DIGITAL SIMULATION	6	1.25	126
INFO 623 Content Representation			
INDEXING	16	2.66	926
THESAURI	12	1.81	278
VOCABULARY	11	1.49	177
SUBJECT INDEX TERMS	6	1.41	158
ABSTRACTING	10	1.01	90
INFO 624 Information Retrieval Systems			
INSYS 300 Information Retrieval Systems			
INFORMATION RETRIEVAL	17	3.99	6010
INFORMATION SCIENCE	15	3.50	3040
INFORMATION RETRIEVAL SYSTEMS	8	3.30	2287
INFORMATION RETRIEVAL SYSTEM EVALUATION	7	1.73	250
INFO 625 Cognition and Information Retrieval			
INFORMATION RETRIEVAL	17	3.99	6010
INFORMATION SCIENCE	15	3.50	3040
USER SATISFACTION (INFORMATION)	7	1.84	291
RELEVANCE (INFORMATION RETRIEVAL)	7	1.65	222
INFORMATION SEEKING	7	1.58	202
INFO 626 Language Processing			
INSYS 360 Language Processing			
NATURAL LANGUAGES	8	1.93	330
LINGUISTICS	8	1.44	166
Info 627 Rquirements Engineering and Management			
INFORMATION NEEDS	15	3.10	1720
INFORMATION USE	7	2.03	382
INFORMATION UTILIZATION	7	1.36	147
INFO 628 Information Systems Implementation			
INSYS 365 Database Administration			
DP MANAGEMENT	9	2.39	628
GOVERNMENT DATA PROCESSING	9	1.92	324
QUALITY CONTROL	13	1.57	199
DP INDUSTRY	8	1.54	190
SECURITY OF DATA	8	1.50	180
ADMINISTRATIVE DATA PROCESSING	7	1.41	158
BUSINESS DATA PROCESSING	5	1.33	142
EDUCATIONAL ADMINISTRATIVE DATA PROCESSING	7	1.25	126
FINANCIAL DATA PROCESSING	7	1.12	105
INFO 630 Evaluation of Information Systems			
EVALUATION METHODS	9	1.94	334
PERFORMANCE EVALUATION	6	1.58	202
EVALUATION CRITERIA	8	1.54	189
INFO 632 Information Services Design and Evaluation			
INFORMATION SERVICES	16	4.01	6208
EVALUATION METHODS	9	1.94	334
MARKETING	16	1.82	283
TREND ANALYSIS	8	1.01	90
INFO 635 Scholarly and Professional Communication			
INDUSTRIAL PROPERTY	8	1.90	316
CHEMISTRY COMPUTING	6	1.48	174
RESEARCH INITIATIVES	8	1.42	161
TECHNOLOGY TRANSFER	10	1.31	137
SCIENTIFIC AND TECHNICAL INFORMATION	8	1.30	136
RESEARCH AND DEVELOPMENT MANAGEMENT	8	1.17	113
BIOLOGY COMPUTING	5	1.08	100
RESEARCH AND DEVELOPMENT	8	1.06	96
INFO 636 Software Engineering Process I			
INFO 637 Software Engineering Process II			
INFO 638 Software Project Management			
SOFTWARE PACKAGES	8	2.19	475
SOFTWARE ENGINEERING	8	1.94	334
SOFTWARE LIBRARIES	1	1.61	210
SOFTWARE REUSABILITY	5	1.25	126

component of which is a small database in which IST courses are subject-indexed so as to bring out both similarities and differences. Analysis of this database could bring out commonalities of theme that emerge when courses share descriptors. Good modules would take a unifying theme such as Management, Communication Skills, or Ethics and reinforce it through multiple different learning experiences; bad modules would simply rehash the same content repeatedly. (Like many colleges employing numerous adjunct faculty, IST is concerned with preventing the "rehash" phenomenon.)

Table 9 illustrates how curricular gaps can be identified. It contains listings of one or more strong-signal terms that apparently do not correspond to anything in the curricula IST now offers. Possible future courses (labeled INFO ???) would incorporate these terms in their subject matter — courses, for example, in Policy Analysis, Medical Informatics, Telecommunications, and Electronic and Optical Publishing. This way of making gaps explicit is a feature of descriptor-based domain analysis that is almost guaranteed to interest faculty and administrators.

A handful of strong-signal terms, e.g., STUDENT ATTITUDES, could not readily be placed. However, Table 10 presents a residual set that nicely suggests the strategies and concerns of IST as a whole — for example, COMPUTER SCIENCE EDUCATION, LIBRARY EDUCATION, and COMPUTER-ASSISTED INSTRUCTION. The Kellogg project (and the present paper) attest to IST's interest in CURRICULUM DEVELOPMENT. The college has long offered TRAINING to certain clienteles (e.g., IBM employees; Philadelphia public school teachers) through minicourses; the development of TRAINING METHODS is naturally a part of this. In common with many of its peers, IST has also ventured into DISTANCE EDUCATION. In 1996, for example, aided by a $1.4 million grant

Table 9
Virtual Curriculum terms suggesting new courses

INFO ??? Standards			
STANDARDS	18	2.42	657
STANDARDISATION	9	1.29	134
GUIDELINES	7	1.50	181
LIBRARY STANDARDS	4	1.03	92
INFO ??? Policy Analysis			
GOVERNMENT POLICIES	8	2.04	384
ECONOMICS [of Information]	10	1.56	195
STATE PROGRAMS	7	1.52	185
GOVERNMENT ROLE	6	1.43	164
PUBLIC POLICY	6	1.21	119
NATIONAL PROGRAMS	6	1.04	94
INFO ??? Medical Informatics			
MEDICAL ADMINISTRATIVE DATA			
MEDICAL COMPUTING	8	2.17	466
PROCESSING	9	2.14	442
HEALTH CARE	9	1.62	212
INFO ??? Computer Graphics and Imaging			
COMPUTER GRAPHICS	12	1.78	266
COMPUTERISED PICTURE PROCESSING	4	1.02	91
IMAGE CODING	1	1.04	94
INFO ??? Computing for Handicapped			
DISABILITIES	8	1.54	189
HANDICAPPED AIDS	6	1.08	100
INFO ??? Records Management			
DOCUMENT IMAGE PROCESSING	5	1.75	257
RECORDS MANAGEMENT	13	1.26	129
ARCHIVES	6	1.03	93
INFO ??? Telecommunications			
TELECOMMUNICATIONS	9	2.31	562
TELECOMMUNICATION SERVICES	5	1.90	317
TELECOMMUNICATION NETWORK MANAGEMENT	4	1.43	163
TELECOMMUNICATIONS COMPUTING	6	1.33	142
TELECOMMUNICATION TRAFFIC	1	1.24	125
TELECOMMUNICATION NETWORKS	4	1.18	114
TELECOMMUNICATION STANDARDS	3	1.01	90
INFO ??? Electronic, Print, and Optical Publishing			
CD-ROMS	8	2.82	1165
ELECTRONIC PUBLISHING	16	2.39	629
OPTICAL DATA DISKS	9	2.20	480
PUBLISHING INDUSTRY	8	1.12	105
COPYRIGHTS	5	1.07	98
OPTICAL DISC STORAGE	6	1.05	95
INFORMATION INDUSTRY	7	1.03	93
OPTICAL PUBLISHING	4	1.03	93
INFO ??? Information Use Studies			
USER NEEDS (INFORMATION)	9	3.13	1793
USERS (INFORMATION)	9	2.44	679
COMMUNICATION (THOUGHT TRANSFER)	8	1.14	108
INFO ??? Computers in Education			
HIGHER EDUCATION	9	3.36	2482
ELEMENTARY SECONDARY EDUCATION	9	2.72	1007
EDUCATION	10	2.41	654
EDUCATIONAL TECHNOLOGY	10	2.19	479
EDUCATIONAL COMPUTING	8	1.84	289
EDUCATIONAL COURSES	8	1.72	246
SECONDARY EDUCATION	9	1.46	171
COLLEGE ADMINISTRATION	5	1.45	168
EDUCATIONAL CHANGE	8	1.45	167
POSTSECONDARY EDUCATION	9	1.40	157
ELEMENTARY EDUCATION	6	1.33	142
EDUCATIONAL RESEARCH	7	1.30	136
TWO YEAR COLLEGES	7	1.20	117
COMMUNITY COLLEGES	6	1.09	101
COLLEGE FACULTY	8	1.04	94
COLLEGE STUDENTS	7	1.00	89
INFO ??? Geographic Information Systems			
GEOGRAPHIC INFORMATION SYSTEMS	7	1.25	127

Table 10
Global terms suggesting IST

Computer aided instruction	9	2.57	811
Training	16	2.43	673
Computer assisted instruction	9	2.26	529
Computer science education	10	2.02	375
Asynchronous transfer mode	2	1.96	343
Teaching	7	1.89	314
Library education	6	1.66	225
Distance education	8	1.64	219
Curriculum development	9	1.61	211
Teaching methods	7	1.53	188
Courseware	7	1.46	170
Library associations	5	1.39	154
Conferences	6	1.15	109
Training methods	7	1.11	104

from the Sloan Foundation, it launched its Asynchronous Learning Network, whereby the MSIS degree is being offered to geographically dispersed employees of large corporations, an example of new COURSEWARE delivered in ASYNCHRONOUS TRANSFER MODE.

To summarize, these apposite descriptors were generated automatically, in a retrieval designed to draw out the implications of a college's main themes. While the present context highlights their relevance to IST, they are, of course, relevant to trends in many colleges and departments, and doubtless would appear in retrievals other than the one shown here. Those who would try the methodology can be advised to trust in the interconnectedness of terms.

4. Conclusion

Several years have passed since the ERIC and INSPEC descriptors of this study were obtained. In the meantime IST has significantly revised its curriculum and planned new curricular initiatives. Thus, there is now some perspective on how the descriptor-based analysis affected the revision. It is possible to state that:

- The analysis contributed to a positive climate for change. Some revisions would have been made in any case, but the descriptor displays helped to crystallize previously vague ideas for reform. They furnished corroborative evidence for innovation where evidence had been lacking.
- No one's creativity has been inhibited by the analysis. No administrator or faculty member has argued that teachers cannot try something new because "it is not warranted by the literature." (That fear was actually raised by one referee, as if descriptors might be given prescriptive force. There are better things to be paranoid about.)
- As might be expected, IST faculty members consulted the full Outline mainly to see that their own interests were adequately represented. None has reported disappointment on this

score (although it has been noted that descriptors related to reference librarianship are curiously sparse).

- The Outline and the Virtual Curriculum provided terms or concepts for a list of competencies that, in the judgment of IST faculty, information professionals should possess. These competencies were named and prioritized for some 19 categories of professionals — occupational "niches" such as information retrieval specialist, multimedia developer, and systems librarian (cf. Childers, 1998; White, 1998).
- Faculty have responded to some of the curricular gaps identified by the analysis. New courses (and possible degree programs) in electronic publishing, knowledge management, and science informatics are under consideration. As of 1999, the core requirements for the MS in library and information science have been augmented with additional work in online retrieval, the Internet, professional and social issues, research and evaluation methods, and information service design, as suggested by the analysis. The core courses for the MSIS were recast to include more of the object-oriented paradigm.
- While the Drexel Curriculum Workbench is still under development, a prototype incorporating the indexing of courses as seen in Table 8 has been built and placed on the World Wide Web (White, 1998). Descriptions of the courses can be retrieved both by the ERIC or INSPEC terms assigned to them and by the natural language of the descriptions (i.e., keywords). Integrated with this presentation are pages that permit retrievals from a matrix that sets forth the 19 niches and the competencies associated with them, and a small database of employers' want-ads from the Philadelphia area.
- The 1995 retrieval furnishes a baseline set of descriptors against which a future retrieval of descriptors can be compared. It will be interesting to see which topics increase or decrease in prominence, as measured by ranked term-weights. Trend analyses of the data are built into the methodology for curriculum evaluation and revision that IST promised Kellogg. This methodology defines a knowledge base that is already partly covered by the college and that is open to further coverage by new courses or subcourses. It directly links the college, as delineated by the nine master terms, with the literatures on which an IST education rests and that are the societal warrants for its curricula.

References

Allen, B. S. et al. (1993). Computer-based mapping for curriculum development. In *Proceedings of Selected Research and Development Presentations. 15th Convention of the Association for Educational Communications and Technology, January 13–17*. New Orleans, LA: Sponsored by the Research and Theory Division.

Buckenham, Sandy et al. (1986). An application of computers to curriculum review and planning. *Journal of Medical Education*, *61*(1), 41–45.

Childers, Thomas (1998). Continuous, market-driven, structured curriculum change: an unfinished odyssey. (Unpublished paper).

Coulter, Neal, Monarch, Ira, & Suresh, Konda (1998). Software engineering as seen through its research literature: a study in co-word analysis. *Journal of the American Society for Information Science*, *49*(13), 1206–1223.

Dameru, Fred J. (1990). Evaluating computer-generated domain-oriented vocabularies. *Information Processing and Management*, *26*(6), 791–801.

Dey, S., & Mand, L. R. (1992). Current trends in computer science curriculum: a survey of four-year programs. *SIGCSE Bulletin*, *24*(1), 9–14.

Diamond, Robert M. (1989). *Designing and improving courses and curricula in higher education: a systematic approach*. San Francisco, CA: Jossey-Bass.

Edmondson, Katherine M. (1993). Concept mapping for the development of medical curricula. *Journal of Research in Science Teaching*, *32*(7), 777–793.

English, Fenwick W. (1980). Curriculum mapping. *Educational Leadership*, *37*(7), 558–559.

Gaff, Jerry G., & James, L. Ratcliff, eds. (1996). *Handbook of the undergraduate curriculum: a comprehensive guide to purposes, structures, practices, and change*. San Francisco, CA: Jossey-Bass.

Gambill, S., & Jackson, W. (1992). Applicability of MIS curriculums to the business environment: an examination of business criticism and an academic response. *Journal of Computer Information Systems*, *33*(1), 13–17.

Glass, R. L. (1992). A comparative analysis of the topic areas of computer science, software engineering, and information systems. *Journal of Systems and Software*, *19*(3), 277–289.

IS '97: Model curriculum and guidelines for undergraduate degree programs in information systems. (1997). Association for Computing Machinery, Association for Information Systems, Association of Information Technology Professionals.

Jackson, Philip W., ed. (1992). *Handbook of research on curriculum: a project of the American Educational Research Association*. New York: Macmillan.

Kanter, S. L., Miller, R. A., Tan, M., & Schwartz, J. (1994). Using POSTDOC to recognize biomedical concepts in medical school curricular documents. *Bulletin of the Medical Library Association*, *82*(3), 283–287.

Kellogg Project (1998). Drexel's curriculum for information and computing professions, sponsored by the W.K. Kellogg Foundation under HRISM: Human Resources for Information Systems Management (PR0007694). Described at http://www.cis.drexel.edu/research/kellogg/index.html.

Lancaster, F. W., & Loescher, Jane (1994). The corporate library and issues management. *Library Trends*, *43*(2), 159–169.

Laribee, J. F. (1992). Building a stronger IRM curriculum: views from IS managers and educators. *Information Systems Management*, *9*(2), 22–28.

Lee, Denis M. S., Trauth, Eileen M., & Farwell, Douglas (1995). Critical skills and knowledge requirements of IS professionals: a joint academic/industry investigation. *MIS Quarterly*, *19*(3), 313–340.

Litecky, C., & Arnett, K. (1992). Job skill advertisements and the MIS curriculum: a market-oriented approach. *Interface: The Computer Education Quarterly*, *14*(4), 45–52.

Reitsch, A. G., & Nelson, F. E. (1990). The use of industry and student perceptions in the redesign of an MIS curriculum. *Interface: The Computer Education Quarterly*, *12*(3), 6–13.

Richards, R. M., & Sanford, C. C. (1992). An evolutionary change in the information systems curriculum at the University of North Texas. *Computers and Education*, *19*(3), 219–228.

Rosen, Robert L. et al. (1992). Using a database to analyze core basic science content in a problem-based curriculum. *Academic Medicine*, *67*(8), 535–538.

Sayles, Jeremy W. (1985). Course information analysis: foundation for creative library support. *Journal of Academic Librarianship*, *10*(6), 343–345.

Septon, Richard D. (1983). A computer-assisted method for analyzing curriculum content. *Journal of Optometric Education*, *8*, 24–29.

Spicer, J. (1994). Curriculum planning: an employer in-depth personal-interview approach. *Journal of Computer Information Systems*, *34*(3), 34–40.

Toombs, William, & Tierney, William G. (1991). *Meeting the mandate; renewing the college and departmental curriculum*. ASHE-ERIC Higher Education Report No. 6. Washington, DC: The George Washington University, School of Education and Human Development.

Trauth, Eileen M., Farwell, Douglas W., & Lee, Denis (1993). The IS expectation gap: industry expectations versus academic preparation. *MIS Quarterly*, *17*(3), 293–307.

White, Howard D. (1998). Drexel curriculum workbench. Available at http://white.cis.drexel.edu/analyzers.html.

Womble, Myra N. (1993). Assessment of competencies for computer information systems curricula. *Delta Pi Epsilon Journal*, *35*(2), 69–85.

Wright, Carol, & Larson, Mary Ellen (1990). Basic information access skills: curriculum design using a matrix approach. *Research Strategies*, *8*(3), 104–115.

Zhu, Donghua, & Porter, Alan L. (1998). *Technology opportunities analysis: illustrated for the case of "knowledge discovery in databases and data mining*." Atlanta, GA: Technology Policy and Assessment Center, Georgia Institute of Technology. Available at http://tpac.gcatt.gatech.edu/~donghua/toakdddm/toakdddm.pdf.