Information Science in a Ph.D. Computer Science Program

G. SALTON
Cornell University, Ithaca, New York*

This report contains recommendations on a sample course curriculum in the general area of information organization and information system design in a Ph.D. Computer Science Program. The subject area is first briefly described, followed by a listing of some desirable graduate-level courses. Suitable bibliographies are appended.

KEY WORDS AND PHRASES: course curriculum, graduate courses, university courses, computer science curriculum, information science, information organization, information retrieval, data retrieval, language analysis, information processing
CR CATEGORIES: 1.52, 3.42, 3.70

1. Information Science

Information science deals with the problems of information handling in the modern world. It is concerned with the structure and properties of information items, the techniques for information handling, the characteristics of information processing devices, and the design and operation of information handling systems.

The main information handling operations are the *production* and *collection* of information items, starting with their initial generation; the *analysis* and *transformation* of information, including editing, printing, and/or publishing, as well as indexing, cataloging, abstracting, and summarizing of information; the *organization* and *storage* of information, including classification and deposit in data banks, libraries, and information repositories; the *encoding* and *transmission* of information from originating points to desired destinations; and finally, the *retrieval* of information in response to information requests, and the *dissemination* of the retrieved data.

One may ask why the field of information science should be of interest to students in Computer Science and what, if anything, computer scientists would contribute to this area. A short and somewhat simplified answer would cite the role of computers as information processing devices and express the hope that computers could eventually be used to carry out all information processing and dissemination tasks selectively, rapidly, and cheaply. A detailed response could include evidence of advances in our understanding of the underlying structures of the information which computing devices are called upon to process. Furthermore, recent hardware and software developments could be cited leading to the experimental implementation of so-called computer utilities; such utilities may eventually offer computing and data processing services on demand to large user classes who may be situated in locations remote from the central equipment. The experiments recently conducted with such time-sharing systems lead to the conclusion that the advent of mechanized networks of data banks and information centers, accessible from many points by many users, may not be so far in the future as might have otherwise been expected.

The information science curriculum should then lead to an examination of the principal technical and intellectual problems arising in information science, and to a description of various proposed procedures for processing information by automatic or semi-automatic methods. The field may be subdivided into three main parts:

(a) the study of information and *data structures*, including in particular the structure of programming and natural languages;

(b) the study of *computer organizations* and operating systems of the kind providing simultaneous service to several types of user classes;

(c) the study of automatic *text processing systems* including text analysis and classification systems, as well as systems for information retrieval and dissemination.

The area of *data structures* should include the study of abstract structures such as vectors, matrices, trees and graphs, as well as the examination of methods for representing, storing, and manipulating these structures by

* Department of Computer Science

This is the fourth in a series of solicited articles on research and teaching areas that might be involved in doctoral programs. The purpose of the series is to provide some guidelines for what constitutes a "good" doctoral program. These articles contain the opinions and recommendations of experts in the subject areas. However, unlike "Curriculum 68," they reflect neither the views of official ACM nor the deliberations of a committee. The series was initiated in the November 1968 issue of CACM.—P. C.

automatic methods. In addition, the main language systems useful in information system design should be described, including programming and natural languages, with particular emphasis on methods for the determination and representation of the syntactic and semantic structure of these languages.

The *computer organization* area includes the study of the multiuser environment where several console initiated jobs—such as information requests received from a user population—would be handled simultaneously with standard file maintenance and other batch processing work. Methods should be covered for organizing and accessing the files in storage, for allocating memory space as required, for insuring that required programs and data be "rolled" in and out of the main store as needed, and that the integrity of the information be protected against outside interference, and for seeing to it that users are serviced in accordance with some established priority schedule. The various file organization and traffic control procedures can be related to their use in several kinds of information processing systems.

The *information systems* area includes an examination of automatic methods and procedures useful in information processing systems such as automated inquiry systems, data collection systems, conversational computing systems, information distribution systems, interactive graphic systems, and remote document production systems. On the assumption that much of the information to be processed consists of natural language data, methods can be covered for the automatic content analysis of written texts, including dictionary construction and look-up procedures, statistical and syntactic language analysis, text storage and search procedures, and retrieval methodology.

Special attention can be given to procedures useful in a time-sharing environment, including file organizations leading to fast searching and rapid responses, and interactive search methods in which queries are refined following negotiations between user and system. Attention must also be given to the design of retrieval languages, and to procedures for the evaluation of retrieval effectiveness. Finally, various special purpose devices can be examined which are useful in automatic text processing, including automatic character recognition, microstorage devices, automatic typesetting and photocomposing, and communications equipment.

2. Course Schedule

A set of six graduate-level courses is proposed which would provide a suitable introduction to the information science field. These courses, numbered 1 to 6, can be considered as three one-year sequences, consisting respectively of 1 and 2, 3 and 4, 5 and 6. The courses could be taught independently, although course 1 might be used best as a general introduction, and a prerequisite to the others. It is assumed that students taking these courses are already familiar with standard computer organizations and have had experience with machine language programming. Specifically, the basic courses numbered B1, B2, and B3 in the ACM Curriculum Committee Report (Introduction to Computing, Computers and Programming, and Introduction to Discrete Structures) are assumed as prerequisites [1]. Furthermore, certain other special purpose undergraduate courses might also usefully be taken to provide an introduction to the field, such as for example the course titled "Introduction to Computational Linguistics" recently described by Kuno and Oettinger [2].

The six proposed courses provide a thorough coverage of language analysis and text processing. Other courses in related areas could be used to supplement the proposed curriculum, including courses in computer graphics, system simulation, or heuristic programming and artificial intelligence. Furthermore, a coverage of some of the more conventional subjects would have to be supplied if the proposed curriculum were to be used in certain professional schools. For example, standard library classification systems may be covered separately in library schools, or legal documentation in law schools.

The bibliography which is appended provides a selective coverage, first of some general texts for the whole area, and then separately for each of the six proposed courses. Other supplementary references will suggest themselves in each case to students and instructors as the study of the various topics is undertaken in greater depth.

Course 1. Data Structures and Information Organization

Review of elements of set theory; Vectors and vector matching operations; Matrix operations and representation: row and column lists, row chained and column chained matrices, sparse matrices.

Tree operations and representation: right and left list matrix, chain list matrix, partitioning and transformation to optimal form.

Lattices and graphs: list matrix, level indication, connection matrix, graph matching methods.

Representation of structured operands: linear and chained representation, forward, backward, and block chaining, partitioned representation.

Storage pools: dimension and address ordered, marked pools, pushdown stores, last-in-first-out and first-in-first-out disciplines.

Search and sorting methods: sequential scanning, controlled scanning, key-address transformations, overflow procedures, chaining methods, serial sorting, internal sorting, sort evaluation.

Prerequisite: Knowledge of machine language programming and possibly a course in discrete algebra.

This course is an introduction to structured operands as well as their representation and manipulation. It describes the constructs of interest in automatic information processing systems, and introduces the principal types of required operations.

Course 2. Time-Sharing Computer Organization

Introduction to multiprogramming environment: user environment, supervisory system.

Addressing techniques: relocation, base registers, two-dimensional addressing, virtual memory, pages and segments.

File organization: file access and protection mechanisms, page and segment management, page turning methods, search methods.

Traffic control: priority control, process switching, scheduling algorithms.

Program intercommunication: foreground and background work, storage management, storage allocation methods.

Prerequisite: Knowledge of machine language programming and standard computer organization.

This is an introduction to the multiprogramming computer environment where real-time (foreground) processing takes place simultaneously with batch processing (background) work. The course might best be scheduled following course 1.

Course 3. Language Structure and Syntactic Analysis

Syntax and semantics of arithmetic expressions: Backus Normal Form representation, translation from parenthesized to parenthesis-free notation, use of pushdown store for translation.

Precedence grammars: operator precedence, syntactic analysis of precedence grammars.

Context-free grammars: recognition methods, top-down and bottom-up algorithms, normal form for context-free grammars.

Structure of natural language: concepts and relations, triplet structure, syntax and semantics.

Context-sensitive and transformational grammars: use in analyzing natural language, recognition methods, efficiency criteria.

Prerequisite: Course 1

This course describes various models of the structure of natural languages, and methods for the automatic syntactic analysis of natural languages.

Course 4. Text Analysis and Automatic Classification

Information models: entities and relations, term vectors and vector mappings, tree and graph models, matrix and lattice models, general diagram model.

Text analysis: language characteristics, word relations and relational indicators, transformation to formal entities, distance concept, text comparisons and transformations.

Thesaurus operations: semi-automatic and fully-automatic construction, thesaurus set-up operations, thesaurus look-up methods, suffix-stem operations, phrase detection, hierarchy set-up and look-up, hierarchical expansion operations.

Statistical operations: term-document mapping, property vector operations, statistical associations, term and document similarity, linear associative retrieval.

Automatic classification: eigenvalue analysis, clustering methods, classification vector, use in searching and retrieval.

Syntactic operations: automatic content analysis, phrase processing, syntactic tree matching.

Basic retrieval process: keyword matching, direct and inverted files, combined file system, vector merging and matching, vector matching of cluster vectors, iterative search methods.

Prerequisite: Course 1 and optionally course 3.

This is an introduction to automatic text analysis methods of the kind useful in automatic text processing and retrieval systems.

Course 5. Information Retrieval System Design

Information dissemination process: composition and typesetting, classification and analysis, abstracting and indexing, retrieving and evaluation.

Information centers: document depots and libraries, abstracting and indexing services, information analysis.

Input operations: input conversion and validation, encoding and editing, numeric and alphabetic codes, special-purpose codes, superimposed coding, query statement formulation.

File organization: bucket assignment, disk and tapestrip organization.

Search strategy: association coefficients, search and look-up operations, multi-level search, iterative negotiated search, adaptive retrieval strategy.

Output operations: dictionary and text display, relevance feedback, query modification.

Language design: query languages, domain and semantic properties, conversational system.

Retrieval evaluation: environment and parameters, recall and precision, determination of recall value, macro and micro-evaluation, pair-wise comparisons, significance computations.

Prerequisite: Course 1 and possibly course 2; also possibly an introductory statistics course.

This covers the design of automatic retrieval systems and gives examples of the principal search and retrieval operations (exclusive of automatic text analysis).

Course 6. Automatic Text Processing Systems

Special purpose equipment: character recognition, optical film scanning, typesetting and photocomposing, storage equipment, micro-storage equipment, teletypewriters and display equipment, alphanumeric and graphical displays, communications and transmission, card and film equipment.

Auxiliary information services: text editing systems, automatic publication systems, paraphrasing systems, index and glossary production, citation indexing, selective information dissemination, concordance preparation.

Abstracting and indexing: term-oriented indexes, document oriented indexes, automatic extracting, term and sentence significance, automatic abstracting.

Data-base retrieval: organization of data base, semantic interpretation, extension of data base, automatic deduction.

On-line retrieval: conversational systems, console operations, tutorial sequences.

Applications: information networks, management information systems, interlibrary systems.

Social problems: file ownership, copyright problems, privacy protection.

Prerequisite: Course 1 and possibly course 5

This covers the design and operations of data-base retrieval systems and information services other than retrieval.

SELECTIVE BIBLIOGRAPHY IN AUTOMATIC INFORMATION PROCESSING

A. General Texts

BECKER, J., AND HAYES, R. M. *Information Storage and Retrieval Tools, Elements, Theories.* Wiley, New York, 1963, 448 pp.

BOURNE, C. P. *Methods in Information Handling.* Wiley, New York, 1963.

CUADRA, C. A. (Ed.) *Annual Review of Information Science and Technology, Vols. 1, 2.* Interscience, New York, 1966, 1967; *Vol. 3.* Encyclopedia Britannica, Chicago, 1968, pp. 389, 484, and 457.

FAIRTHORNE, R. A. *Towards Information Retrieval.* Butterworth, London, 1961, 211 pp.

KENT, A. *Textbook on Mechanized Information Retrieval.* Interscience, New York, 1963.

MEADOW, C. T. *The Analysis of Information Systems.* Wiley, New York, 1967, 301 pp.

SALTON, G. *Automatic Information Organization and Retrieval.* McGraw-Hill, New York, 1968, 514 pp.

SHARPE, J. R. *Some Fundamentals of Information Retrieval.* André Deutsch, London, 1965, 224 pp.

VICKERY, B. C. *On Retrieval System Theory.* Butterworth, London, 1961.

WILLIAMS, W. F. Principles of Automated Information Retrieval. Taplinger, New York, 1965.

B. Retrieval Environment

BROWNSON, H. L. Research on handling scientific information. *Science 132*, 3444 (Dec. 30, 1960), 1922–1931.

CARTER, L. F. National document-handling systems in science and technology. *Science. 154*, 3754 (Dec. 9, 1966), 1299–1303.

DE SOLLA PRICE, D. J. Networks of scientific papers. *Science 149*, 3683 (July 30, 1965), 510–515.

KASHER, A. Data retrieval by computer—a critical survey. In Kochen, M. (Ed.), *The Growth of Knowledge*, Wiley, New York, 1967, pp. 292–324.

LICKLIDER, J. C. R. *Libraries of the Future*. MIT Press, Cambridge, Mass., 1965.

LIPETZ, B. A. Information storage and retrieval. *Sci. Amer. 215*, 3 (Sept. 1966), 224–238.

National Science Foundation: Current research and development in scientific documentation No. 14, NSF Rep. 66–17, 1966, 662 pp.

OVERHAGE, C. F. J. Plans for project intrex. *Science 152*, 3725 (May 20, 1966), 1032–1036.

SALTON, G. Progress in automatic information retrieval. *IEEE Spectrum 2*, 8, (Aug. 1965), 90–103.

SEE, R. Machine aided translation and information retrieval. In Kent, A., et al. (Eds.), *Electronic Handling of Information: Testing and Evaluation*, Thompson Book Co., Washington, D.C., 1967, pp. 89–108.

SIMMONS, R. F. Automated language processing. In Cuadra, C. (Ed.), *Annual Review of Information Science and Technology, Vol. 1*, Interscience, New York, 1966, pp. 137–170.

SWANSON, D. R. Information retrieval—state of the art, Proc. Western Joint Comput. Conf., May, 1961, Spartan Books, New York, pp. 239–246.

C. Data Structures (Course 1)

BERGE, C. *Theory of Graphs and Its Applications*. Wiley, New York, 1962.

CROS, R. C., GARDIN, J. C., AND LÉVY, F. *L'Automatisation des recherches documentaries—Un modèle général "Le Syntol."* Gauthier-Villars, Paris, 1964, 260 pp.

DAHL, O. J., AND NYGAARD, K. SIMULA—an ALGOL-based simulation language. *Comm. ACM 9*, 9 (Sept. 1966), 671–678.

DE GROLIER, E. A Study of General Categories Applicable to Classification and Coding in Documentation. UNESCO, 1962, 248 pp.

DOYLE, L. B. Semantic road maps for literature searches. *J. ACM 8*, 4 (Oct., 1961), 553–578.

HILLMAN, D. J. Mathematical classification techniques for non-static document collections with particular reference to the problem of relevance. In *International Study Conference on Classification Research*, P. Atherton (Ed.), Munksgaard, Copenhagen, 1965.

HOLT, A. W. Some theorizing on memory structure and information retrieval. *Proc. ACM 18th Natl. Conf.*, 1963, 27 pp.

IVERSON, K. E. *A Programming Languages*. Wiley, New York, © 1962, 286 pp.

KNUTH, D. E. *The Art of Computer Programming, Vol. 1, Fundamental Algorithms*. Addison-Wesley, Reading, Mass., 1968, 634 pp.

LEFKOWITZ, D., AND POWERS, R. V. A list-structured chemical information retrieval system. In Schecter, G. (Ed.), *Information Retrieval—A Critical View*, Thompson Book Co., Washington, D.C., 1967, pp. 109–130.

——, AND PRYWES, N. S. Automatic stratification of information. Proc., AFIPS 1963 Spring Joint Comput. Conf., Vol 23, Spartan Books, New York, pp. 229–240.

MINKER, J., AND SABLE, J. File organization and data management. In Cuadra, C. A. (Ed.), *Annual Review of Information Science and Technology, Vol. 2*, Interscience, New York, 1967, pp. 123–160.

MOOERS, C. N. A mathematical theory of language symbols in retrieval. Proc. Intern, Conf. Sci. Inform., Washington, D.C., 1958, pp. 1327–1364.

PARKER RHODES, A. F., ET AL. A lattice model of syntactic description. Repts. ML 147 and others, Cambridge Lang. Res. Unit, July, 1961.

PRYWES, N. S., AND GRAY, H. J. The multi-list system for real-time storage and retrieval. Proc. IFIP Congr. Munich, 1962.

RICHMOND, P. A., Transformation and organization of information content: classification research, Proc. 1965 FID Congress, 1966, pp. 87–106.

SALTON, G. Data manipulation and programming problems in automatic information retrieval. *Comm. ACM 9*, 3 (Mar. 1966), 204–210.

SHOFFNER, R. M. A technique for organizing large files. *Am. Doc. 13*, 1 (Jan. 1962), 95–103.

STANDISH, T. A. A data definition facility for programming languages. Ph.D. Th. Carnegie Inst. of Tech. Pittsburgh, Pa., 1967, 292 pp.

SUSSENGUTH, E. H., JR. Use of tree structures for processing files. *Comm. ACM 6*, 5 (May 1963), 272–279.

WEGNER, P. *Programming Languages, Information Structures, and Machine Organization*. McGraw-Hill, New York, 1968, 401 pp.

D. Processor Organization (Course 2)

ARDEN, B. W., GALLER, B. A., O'BRIEN, T. C., AND WESTERVELT, F. H. Program and addressing structure in a time sharing environment. *J. ACM 13*, 1 (Jan. 1966), 1–16.

CORBATO, F. J., MERVIN-DAGGETT, M., AND DALEY, R. C. An experimental time sharing system. In Rosen, S. (Ed.), *Programming Systems and Languages*, McGraw-Hill, New York, 1967, pp. 683–698.

——, AND VYSSOTSKY, V. A. Introduction and overview of the Multics system. Proc. AFIPS 1965 Fall Joint Comput. Conf., Vol. 27, Pt. 1, Spartan Books, New York, pp. 185–196; also in Rosen, S. (Ed.), *Programming Systems and Languages*, McGraw-Hill, New York, 1967, pp. 714–730.

DALEY, R. C., AND DENNIS, J. B. Virtual memory, processes, and sharing in MULTICS. *Comm. ACM 11*, 5 (May 1968), 306–312.

——, AND NEUMANN, P. G. A general purpose file system in secondary storage. Proc. AFIPS 1965 Fall Joint Comput. Conf., Vol. 27, Pt. 1, Spartan Books, New York, pp. 213–229.

DENNIS, J. B. Segmentation and the design of multiprogrammed computer systems, *J. ACM 12*, 4 (Oct. 1965), 589–602; also in Rosen, S. (Ed.), *Programming Systems and Languages*, McGraw-Hill, New York, 1967, pp. 699–713.

FALKOFF, A. D., IVERSON, K. E., AND SUSSENGUTH, E. H. A formal description of System/360. *IBM Syst. J., 3*, 3 (1964), 198–262.

FIKES, R. E., LAUER, H. C., AND VAREHA, A. L., JR. Steps toward a general purpose time sharing system using large capacity core storage and TSS/360. Rep., Dept. of Comput. Sci., Carnegie-Mellon U., Pittsburgh, Pa., Mar. 1968.

HELLER, J. Sequencing aspects of multiprogramming. *J. ACM 8*, 3 (July 1961) 426–439.

HELLERMAN, H. *Digital Computer System Principles*. McGraw-Hill, New York, 1967.

KILBURN, T., EDWARDS, D. B. G., LANIGAN, M. J., AND SUMNER, F. H. One Level Storage System. *IRE Trans. EC-11*, 2 (Apr. 1962), 223–235.

——, PAYNE, R. B., AND HOWARTH, D. J. The Atlas supervisor. In Rosen, S. (Ed.), *Programming Systems and Languages*, McGraw-Hill, New York, 1967, pp. 661–682.

LAMPSON, B. W. A scheduling philosophy for multiprocessing systems. *Comm. ACM 11*, 5 (May 1968) 347–360.

MEALY, G. H. Operating systems. In Rosen, S. (Ed.), *Program-*

ming Systems and Languages, McGraw-Hill, New York, 1967, pp. 516–534.

NIELSON, N. R. The simulation of time sharing systems. *Comm. ACM 10*, 7 (July 1967), 397–412.

RANDELL, B., AND KUEHNER, C. J. Dynamic storage allocation systems. *Comm. ACM 11*, 5 (May 1968), 297–306.

ROSEN, S. (Ed.) *Programming Systems and Languages*. McGraw-Hill, New York, 1967, 734 pp.

SALTZER, J. H. Traffic control in a multiplexed computer system. Ph.D. Th., Rep. MAC-TR-30, MIT, Cambridge, Mass., July, 1966, 77 pp.

SCHWARTZ, J. I., COFFMAN, E. G. AND WEISSMAN, C. A general purpose time-sharing system. Proc. AFIPS 1964 Spring Joint Comput. Conf., Vol. 25, Spartan Books, New York, pp. 397–412.

[Various Authors] *Operating System Principles—Proceedings of an ACM Symposium. Comm. ACM. 11*, 5 (May 1968), 295–377.

E. Syntax and Language Structure (Course 3)

BAR-HILLEL, Y. *Language and Information: Selected Essays on Their Theory and Application*. Addison-Wesley, Reading, Mass., 1964, 388 pp.

BAUER, F. L., AND SAMELSON, K. Sequential formula translation. *Comm. ACM, 3*, 2 (Feb. 1960), 76–83.

BOBROW, D. G. Syntactic analysis of English by computer—a survey. Proc. AFIPS 1963 Fall Joint Comput. Conf., Vol 24, Spartan Books, New York, pp. 365–388.

——, AND RAPHAEL, B. A comparison of list processing computer languages. In Rosen, S. (Ed.), *Programming Systems and Languages*, McGraw-Hill, New York, 1967, pp. 490–515.

BROOKER, R. A., AND MORRIS, D. A general translation program for phrase structure grammars. *J. ACM 9*, 1 (Jan. 1962), 1–10.

CHEATHAM, T. E., JR., AND SATTLEY, K. Syntax directed compiling. Proc. AFIPS 1964 Spring Joint Comput. Conf., Vol. 25, Spartan Books, New York, pp. 31–58.

CHOMSKY, N. *Syntactic Structures*. Mouton and Co., The Hague, 1957, 116 pp.

——, AND MILLER, G. A. Introduction to the formal analysis of natural languages. Luce, R. D., Bush, R. R. and Galanter, E. (Eds.), *Handbook of Mathematical Psychology, Vol. 2*, New York, 1963, pp. 269–321.

FARBER, D. J., GRISWOLD, R. E., AND POLONSKY, I. P. The SNOBOL 3 programming language. *Bell Syst. Tech. J. 45*, 6 (July–Aug. 1966).

FELDMAN, J. A., AND GRIES, D. Translator writing systems. Stanford U. Tech. Rept. CS 69, June 1967; also *Comm. ACM, 11*, 2 (Feb. 1968), 77–113.

FLOYD, R. W. Syntactic analysis and operator precedences. *J. ACM 10*, 3, (July 1963), 316–333.

——, The syntax of programming languages—a survey. In Rosen, S. (Ed.), *Programming Systems and Languages*, McGraw-Hill, New York, 1967, pp. 342–361.

GRAHAM, R. M. Bounded context translation. In Rosen, S. (Ed.), *Programming Systems and Languages*, McGraw-Hill, New York, 1967, pp. 184–205.

GREIBACH, S. A new normal-form theorem for context-free phrase structure grammars. *J. ACM 12*, 1 (Jan. 1965), 45–52.

GRIFFITHS, T. V., AND PETRICK, S. R. On the relative efficiencies of context-free grammar recognizers. *Comm. ACM 8*, 5 (May, 1965), 289–300.

HAYS, D. G., (Ed.) *Readings in Automatic Language Processing*, American Elsevier, New York, 1966, 202 pp.

IRONS, E. A Syntax directed compiler for ALGOL-60. *Comm. ACM 4*, 1 (Jan. 1961) 51–55.

KNUTH, D. E. On the translation of languages from left to right. *Inf. Contr. 8*, 6 (Dec. 1965), 607–639.

KUNO, S. Automatic syntactic analysis. Seminar on Computational Linguistics, Pratt, A. W., Roberts, A. H., and Lewis, K. (Eds.), Public Health Service Publ. 1716, Washington, D.C., 1967, pp. 19–41.

——. Computer analysis of natural languages. Proc. of Symposia in Applied Mathematics, vol. 19, (*Mathematical Aspects of Computer Science*), Amer. Math. Soc., 1967, pp. 52–110.

——, AND OETTINGER, A. G. Multiple-path syntactic analyzer. In *Information Processing—1962*, North Holland Pub. Co., Amsterdam, 1963, pp. 306–312.

LAMPSON, B. W. Interactive machine language programming. Proc. AFIPS 1965 Fall Joint Comput. Conf., Vol. 27, Pt. 1, Spartan Books, New York, pp. 473–481.

LAWSON, H. W., JR. PL/I list processing. *Comm. ACM 10*, 6 (June 1967) 358–367.

LIU, C. L., CHANG, G. D., AND MARKS, R. E. Design and Implementation of a table-driven compiler system. Rep. MAC-TR-42, MIT, Cambridge, Mass., July, 1967, 86 pp.

NAUR, P. (Ed.), Revised report on the algorithmic language, ALGOL 60. *Comm. ACM 6*, 1 (Jan. 1963), 1–17.

OETTINGER, A. G. Computational linguistics, Pt. II. *Amer. Math. Mon. 72*, 2, 1965.

RAPHAEL, B. The structure of programming languages. *Comm. ACM 9*, 2 (Feb. 1966), 67–71.

ROSEN, S. (Ed.) *Programming Systems and Languages*. McGraw-Hill, New York, 1967, 734 pp.

UNGER, S. H. A Global Parser for Context-free Phrase Structure Grammars. *Comm. ACM 11*, 4 (Apr. 1968), 240–247.

WEIZENBAUM, J. Symmetric list processor. *Comm. ACM 6*, 9 (Sept. 1963), 524–544.

WIRTH, N., AND WEBER, H. Euler—A generalization of ALGOL and its formal definition. *Comm. ACM 9*, 1 and 2 (Jan. and Feb., 1965), 13–23 and 89–100.

YNGVE, V. H. A model and a hypothesis for language structure. Proc. Am. Phil. Soc. *104*, 5 (1960).

F. Text Analysis (Course 4)

BONNER, R. E. On some clustering techniques. *IBM J. Res. Develop. 8*, 1 (Jan. 1964), 22–32.

BORKO, H., AND BERNICK, M. D. Automatic document classification, *J. ACM 10*, 2 (Apr. 1963), 151–162.

CHEATHAM, T. E., AND WARSHALL, S. Translation of retrieval requests couched in a semi-formal English-like language. *Comm. ACM 5*, 1 (Jan. 1962), 34–39.

CLIMENSON, W. D., HARDWICK, N. H., AND JACOBSON, S. N. Automatic syntax analysis in machine indexing and abstracting. Third Institute on Information Retrieval, American U., Washington, D. C., 1961; also *Amer. Doc. 12*, 3 (July, 1961), 178–183.

COYAUD, M. Une méthode sémantique d'indexation de documents. T. A. Informations, no. 2, Editions Klincksieck, Paris, 1966.

DENNIS, S. F. The design and testing of a fully automatic index-searching system for documents consisting of expository text. In Schecter, G. (Ed.), *Information Retrieval—A Critical View*, Thompson Book Co., Washington, D. C., 1967, pp. 67–94.

DOYLE, L. B. Indexing and abstracting by association. *Amer. Doc. 13*, 4 (Oct. 1962), 378–390.

——. The microstatistics of text. *Inform. Storage Retrieval 1*, 4 (Nov. 1963), 189–214.

EDMUNDSON, H. P. Problems in automatic abstracting. *Comm. ACM 7*, 4 (Apr. 1964), 259–263.

JONES, P. E., GIULIANO, V. E., AND CURTICE, R. M. Papers on automated language processing: vol. 1, Selected collection statistics; vol. 2, Linear models for associative retrieval; vol. 3, Development of string indexing techniques. Rep. ESD-TR-67-202, Arthur D. Little, Inc., Feb., 1967, 121, 215, and 137 pp.

KLEIN, S., AND SIMMONS, R. F. A computational approach to grammatical coding of english words. *J. ACM 10*, 3 (July 1963), 334–347.

LUHN, H. P. A statistical approach to mechanized encoding and

searching of literary information. *IBM J. Res. Develop. 1*, 4 (Oct. 1957), 309–317.

——. The automatic creation of literature abstracts. *IBM J. Res. Develop. 2*, 2 (1958), 159–165.

MARON, M. E., AND KUHNS, J. L. On relevance, probabilistic indexing and information retrieval. *J. ACM 7*, 3 (July 1960), 216–244; also automatic indexing: and experimental inquiry. *J. ACM 8*, 3 (July 1961), 404–417.

MELTON, J. S. Automatic language processing for information retrieval—some questions. Proc. 1966 Ann. Meeting, Amer. Doc. Inst., Santa Monica, 1966.

——, REEVES, P. W., AND HESPIN, C. F. Automatic processing of metallurgical abstracts for the purpose of information retrieval. Final Rep. NSF-4, Center for Doc. and Comm. Research, Western Reserve U., Cleveland, Ohio, July, 1967, 68 pp.

NEEDHAM, R. M., AND SPARCK-JONES, K. Keywords and clumps. *J. Doc. 20*, 1 (Mar. 1964), 5–15.

O'CONNOR, J. Mechanized indexing methods and their testing. *J. ACM 11*, 4 (Oct. 1964), 437–449.

OETTINGER, A. G. Automatic syntactic analysis and the push-down store. Proc. Symp. Structure of Language and Its Math. Aspects, vol 12, Amer. Math. Soc., 1961.

ROBINSON, J. J. Automatic parsing and fact retrieval: a comment on grammar, paraphrase and meaning. Memo. RM-4005-PR, Rand Corp., Santa Monica, Calif., Feb. 1964.

SALTON, G. Automatic phrase matching. In Hays, D. (Ed.), *Readings in Automatic Language Processing*. American Elsevier, New York, 1966, pp. 169–188.

SPARCK-JONES, K., AND JACKSON, D. M. Recent approaches to classification and clump-finding at Cambridge Language Research Unit. *Comput. J., 10*, 1, (May 1967).

STEVENS, M. B. Automatic indexing: *a State of the art report.* Nat. Bur. Standards, Monograph 91, 1965, 220 pp.

STEVENS, M. E., GIULIANO, V. E., AND HEILPRIN, L. B. (Eds.). *Statistical association methods for mechanized documentation—Symposium Proceedings*. Misc. Pub. 269, US Dept. of Commerce, Washington, D.C., 1964, 261 pp.

STILES, H. F. The association factor in information retrieval. *J. ACM 8*, 2 (Apr. 1961), 271–279.

SWANSON, D. R. Research procedures for automatic indexing. Third Inst. on Inf. Storage and Retrieval, American U., Washington, D.C., 1961, pp. 281–304.

——. Interrogating a computer in natural language. In *Information Processing—62*, North-Holland Pub. Co., Amsterdam, 1963.

SWETS, J. A. Information retrieval systems. *Science 141*, 3577 (July 19, 1963), 245–250.

WILLIAMS, J. N., JR. A discriminant method for automatically classifying documents. Proc. AFIPS 1963 Fall Joint Comput. Conf., Vol. 24, Spartan Books, New York, pp. 161–166.

YNGVE, V. H. A framework for syntactic translation. *Mach. Transl. 4*, 3 (Dec. 1957), 59–65.

G. Retrieval System Design (Course 5)

BARNETT, M. P., AND RUHSAM, W. M. A natural language programming system for text processing. *IEEE Trans. EWS* Aug. 1968, 45–52.

BERUL, L. H. Survey of equipment developments in the information storage and retrieval field. FID/IFIP Conf. on Mechanized Documentation, Rome, 1967.

BOURNE, C. P. Review of the methodology of information systems design. In *Information Retrieval Workshop*, Spartan Books, New York, 1962.

BROOKS, F. P., JR., AND IVERSON, K. E. *Automatic Data Processing*. Wiley, New York, 1963, 494 pp.

BROWNSON, H. L. Evaluation of document searching systems and procedures. *J. Doc. 21*, 4 (Dec. 1965).

BURNAUGH, H. P. The Bold (Bibiographic On-line Display) System. In Schecter, G. (Ed.), *Information Retrieval—A Critical View*, Thompson Book Co., Washington, D. C., 1967, pp. 53–66.

CLEVERDON, C. W., AND MILLS, J. The testing of index language devices. *Aslib Proc. 15*, 4 (1963), 173–193.

——, ——, AND KEEN, M. Factors Determining the Performance of Indexing Systems, Design, vol. 1; Test Results, vol. 2, Aslib-Cranfield Research Project, Cranfield College of Aeronautics, England, 1966, 377 and 299 pp.

CLIMENSON, W. D. File Organization and Search Techniques, in C. A. Cuadra (Ed.), "Annual Review of Information Science and Technology," vol. 1, Interscience, New York, 1966, pp. 107–136.

CUADRA, C., AND KATTER, R. Experimental studies of relevance judgments, final report: project summary, vol. 1; description of individual studies, vol. 2, System Development Corp. Reps. TM-3520/001–002, June 1967, 129 and 135 pp.

DREW, D. L., SUMMIT, R. K., TANAKA, R. I., AND WHITELY, R. B. An On-line Technical Library Reference Retrieval System, *Am. Doc. 17*, 1 (Jan. 1966), 3–7.

FAIRTHORNE, R. A. Some Basic Comments on Retrieval Testing, *J. Doc. 21*, 4 (Dec., 1965.) 267–270.

FALKOFF, A. D., AND IVERSON, K. E. The APL terminal system: instructions for operation, IBM Res. Rep. Yorktown Heights, N. Y. Nov. 1966.

——. Algorithms for Parallel Search Memories, *J. ACM 9*, 4 (Oct. 1962), 488–511.

FELS, E. M. Evaluation of the Performance of an Information Retrieval System by the Modified Mooers Plan, *Am. Doc. 14*, 1 (Jan. 1963), 28–34.

GARDIN, J. C. Syntol. In Artandi, Susan (Ed.), *Systems for the Intellectual Organization of Information, vol. 2*. Rutgers U. Press, New Brunswick, N.J., 1965, 106 pp.

GHOSH, S. P., ABRAHAM, C. T., AND RAY-CHAUDHURI, D. K. File Organization Schemes Based on Finite Geometries, IBM Res. Center Rep. RC-1459 and RC-1561. Aug, 1965, and Mar, 1966, pp. 1–12.

GIULIANO, V. E., AND JONES, P. E. Study and test of a methodology for laboratory evaluation of message retrieval systems, Rep. ESD-TR-66-405, Arthur D. Little, Inc., Aug. 1966, 183 pp.

HELLERMAN, H. Addressing multidimensional arrays. *Comm. ACM 5*, 4 (Apr. 1962), 205–207.

HENDERSON, M. M. *Bibliography on Evaluation of Information Systems*. Nat. Bur. Standards, Washington, D.C., July, 1965, 35 pp.

HIBBARD, T. N. Some combinatorial properties of certain trees with applications to searching and sorting. *J. ACM 9*, 1 (Jan. 1962), 13–28.

HORTY, J. F. Experience with the application of electronic data processing systems in general law. *MULL, 60D*, (Dec., 1960).

MAGNINO, J. J., JR. IBM technical information retrieval centers —progress and plans. Proc. 1966 Ann. Meeting, Am. Doc. Inst., Santa Monica, Calif., 1966.

MATHEWS, WM. D. The TIP retrieval system at MIT. In Schecter, G. (Ed.), *Information Retrieval—A Critical View*. Thompson Book Co., Washington, D.C., 1967, pp. 95–108.

MOOERS, C. N. TRAC—a procedure-describing language for the reactive typewriter. *Comm. ACM 9*, 3 (Mar. 1966), 215–219.

——. The intensive sample test. *Zator Bull. 132* (Aug. 1959).

O'CONNOR, J. Automatic subject recognition in scientific papers: an empirical study. *J. ACM. 12*, 4 (Oct. 1965), 490–515.

PETERSON, W. W. Addressing for random-access storage. *IBM J. Res. Develop. 1*, 2 (Apr. 1957).

REES, A. Semantic factors, role indicators et alia. *Aslib Proc. 15*, 12 (Dec. 1963), 350–363.

——, AND SARACEVIC, T. The measurability of relevance, Proc.

1966 Ann. Meeting, Am. Doc. Inst., Santa Monica, Calif., 1966.

ROCCHIO, J. J. Document retrieval systems—optimization and evaluation. Doctoral th. Report ISR-10, Harvard Comput. Lab. Harvard U., Cambridge, Mass., Mar. 1966, 162 pp.

ROLLING, L. N. A computer-aided information service for nuclear science and technology. *J. Doc. 22*, 2 (June 1966), 29.

SALTON, G., AND LESK, M. E. The SMART automatic document retrieval system. *Comm. ACM 8*, 6 (June 1965), 391–398.

—— AND ——. Computer evaluation of indexing and text processing. *J. ACM 15*, 1 (Jan. 1968), 8–36.

——, Search and retrieval experiments in real-time information retrieval. *Proc. IFIP Congress-68*, North Holland Pub. Co., Amsterdam, 1969, pp. 221–232.

SCHAY, G., AND RAVEN, N. A method for key to address transformation. *IBM J. Res. Develop. 7*, 2 (Apr. 1963), 121–126.

SUSSENGUTH, E. H., JR. A Graph Theoretic Algorithm for Matching Chemical Structures, doctoral th. Rep. ISR-6, Harvard Comput. Lab., Harvard U., Cambridge, Mass., Apr., 1964 pp. 36–43.

SWANSON, D. R. The evidence underlying the Cranfield results. *Library Quart. 35*, 1 (Jan. 1965), 1–20.

SWETS, J. A. Effectiveness of information retrieval methods. Bolt Beranek and Newman Rept. 1499, Cambridge, Mass., Apr. 1967, 47 pp.

TAGUE, J. Evaluation of statistical association measures. Proc. 1966 Ann. Meeting, Am. Doc. Inst., Santa Monica, Calif., Oct. 1966.

TRITSCHLER, R. J. A Computer integrated system for centralized information dissemination, storage and retrieval. *Aslib Proc. 14*, 12 (Dec. 1962), 473–503.

UHR, L. (Ed.) *Pattern Recognition*. Wiley, New York, 1966.

H. Text Processing Systems (Course 6)

BLACK, F. A deductive question-answering system. Doctoral th. Harvard U., Cambridge, Mass., May, 1964, 172 pp.

BOBROW, D. G. A question answering system for high school algebra word problems. Proc. AFIPS 1964 Fall Joint Comput. Conf., Vol. 26, Spartan Books, New York, pp. 591–614.

——, KAIN, R. T., RAPHAEL, B., AND LICKLIDER, J. C. R. A computer program system to facilitate the study of technical documents. *Am. Doc. 17*, 4 (Oct. 1966), 186–189.

BRANDENBERG, W., FALLON, H. C., HENSLEY, C. B., SAVAGE, T. R., AND SOWARBY, A. J. Selective dissemination of information SDI-2 System. IBM Rep., 1961, 97 pp.

COOPER, W. S. Fact retrieval and deductive question-answering information retrieval systems. *J. ACM 11*, 2 (Apr. 1964), 117–137.

CRAIG, J. A., BEREZNER, S. C., CARNEY, H. C., AND LONGYEAR, C. R. Deacon: Direct English Access and Control. *Proc. AFIPS 1966 Fall Joint Computer Conf.*, Vol. 29, Spartan Books, New York, pp. 365–380.

FEIGENBAUM, E. A. The simulation of verbal learning behavior. In *Computers and Thought*, pp. 297–309.

FISCHER, M. The KWIC index concept: a retrospective view. *Am. Doc. 17*, 2 (Apr. 1966), 57–70.

GARFIELD, E. Science citation index. *Science, 144*, 3619 (May 8, 1964), 649–654.

GLUSHKOV, V. M. Certain questions on the theory of machine self learning. Symp. Artificial Intelligence IFIP Congr. 62, Munich, 1962.

GREEN, B. F., WOLF, A. K., CHOMSKY, C., AND LAUGHERY, K. Baseball: an automatic question-answerer. *Proc. Western Joint Computer Conf.*, 1961, pp. 221–224.

KESSLER, M. M. Bibliographic coupling extended in time: ten case histories. *Information Storage and Retrieval 1*, 4 (Nov. 1963), 169–187.

——. Bibliographic coupling between scientific papers. *Am. Doc. 14*, 1 (Jan. 1963).

KOCHEN, M. C. ABRAHAM, AND WONG, E. Adaptive man-machine concept processing. In *Some Problems in Information Science*. Scarecrow Press, Metuchen, N.J., 1965, 156 pp.

KUNEY, J. H. Computers and Scientific Periodicals, presented at the American University Seminar on Electronics and Automation in Publishing, Washington, D.C., May, 1963, 11 pp.

LEVIEN, R., AND MARON, M. E. Relational data file, Pt. 1, design philosophy; Pt. 2, implementation. In Schecter, G. (Ed.), *Information Retrieval—A Critical View*. Thompson Book Co., Washington, D.C., 1967, pp. 211–242.

LINDSAY, R. K. Inferential memory as the basis of machines which understand natural language. In *Computers and Thought*. pp. 217–233.

LUHN, H. P. Keyword-in-Context for Technical Literature. Rep. RC-127, IBM Advanced Syst. Develop. Div., Aug., 1959, 16 pp.

MARGOLIS, J. Citation indexing and evaluation of scientific papers. *Science 155*, 3767 (Mar. 10, 1967), 1213–1219.

NARASIMHAN, R. Syntax-directed interpretation of classes of pictures. *Comm. ACM 9*, 3 (Mar. 1966), 166–173.

RAPHAEL, B. A computer program which "understands." Proc. AFIPS 1964 Fall Joint Comput. Conf., Vol. 26, Spartan Books, New York, pp. 577–589.

SAGE, C. R. Comprehensive dissemination of current literature. *Am. Doc. 17*, 4 (Oct. 1966), 155–177.

SIMMONS, R. F. Answering English questions by computer: a survey. *Comm. ACM 8*, 1 (Jan. 1965), 53–69.

SLAGLE, J. R. Experiments with a deductive question-answering program. *Comm. ACM 8*, 12 (Dec. 1965), 792–798.

STONE, P. J., AND HUNT, E. B. A computer approach to content analysis: studies using the general inquirer. Proc. AFIPS 1963 Spring Joint Comp. Conf., Vol. 23, Spartan Books, New York, pp. 241–256.

SUTHERLAND, I. E. Sketchpad: A man-machine graphical communication system. Lincoln Lab. Tech. Rep. No. 296, MIT, Lexington, Mass., 1963.

SWANSON, D. R. Interrogating a computer in natural language. In *Information Processing-62*. North Holland Publ. Co., Amsterdam, 1963, pp. 288–291.

THOMPSON, F. B. English for the computer. *Proc. AFIPS 1966 Fall Joint Computer Conf.*, Vol. 29, Spartan Books, New York, pp. 349–356.

VAN DAM, A. Computer-driven displays and their uses in man/machine interaction. In Alt, F. L. (Ed.), *Advances in Computers, Vol. 7*. Acadmic Press, New York, 1966, pp. 239–290.

WEIZENBAUM, J. Eliza—A computer program for the study of natural language communications between man and machine. *Comm. ACM 9*, 1 (Jan. 1966), 36–45.

WESTBROOK, J. H. Identifying Significant Research, *Science 132*, 3435 (Oct. 28, 1960), 1229–1234.

RECEIVED SEPTEMBER, 1968; REVISED SEPTEMBER, 1968

REFERENCES

1. Curriculum 68—Recommendations for academic programs in computer science—a report of the ACM Curriculum Committee on Computer Science. *Comm. ACM 11*, 3 (Mar. 1968), 151–197.

2. KUNO, S., AND OETTINGER, A. G. Computational linguistics in a Ph.D. computer science program. *Comm. ACM 11*, 12 (Dec. 1968), 831–836.