

Annotated Bibliography Relating to Automatic Indexing in Information Retrieval Originated by Mitchell Wyle, Adapted by Edward Fox

Advances in Intelligent Retrieval: INFORMATICS 8. Proceedings of an Aslib/British Computer Society Joint Conference, 16-17 April . Aslib, London, 1985.

The following topics were dealt with: extended relational analysis; supercomputers and nonnumeric processing; an architectural approach to advances in information retrieval; word processing and information retrieval; clustering and nearest neighbour searching; automatic book classification; ranking for free text retrieval systems; an artificial intelligence approach to interactive retrieval of biographical data; a case system processor for the PRECIS indexing language; linguistic methods in information retrieval systems; online elucidation of natural language search statements; problem descriptions and user models; automatic intelligent search sequencing in an online public access catalogue; a catalogue advisory system; and an architecture for integrating data, knowledge and information bases. Abstracts of individual papers can be found under the relevant classifier.

Addis, T. R; Johnson, L. Knowledge for Machines. In *The Fifth Generation Computer Project. State of the Art Report.*, Pergamon Infotech, Ltd., Maidenhead, Berkshire, England, 1983.

Knowledge is examined from both the abstract and technological aspects. An approach to the abstract problem of defining knowledge is developed by distinguishing knowledge structures from meaning structures, and connecting knowledge with justification. The essentials of the three major knowledge representation schemes -- semantic nets, production systems, and frames -- are examined. The conclusion is drawn that research is needed on the role of the user within a complete system. To construct a system with knowledge requires the ability to elicit knowledge and to construct an appropriate abstraction at a level of resolution adequate to deal with the set of expected situations for which the system was designed. New skills will have to be evolved, and a new breed of computer expert ('knowledge engineer') will be required.

Appelrath, H. The extension of data base systems and information retrieval systems to knowledge. *Nachrichten fuer Dokumentation*, 36(1):13-21, 1985.

The aim of this paper is to discuss the two traditional types of information systems, namely database (DB) and information retrieval (IR) systems, with respect to their extension into knowledge base systems. The latter are introduced in chapter 1, followed by a discussion of PROLOG in chapter 2. Building on that, chapter 3 treats the extension of DB and chapter 4 deals with the extension of IR systems into knowledge based systems.

Azmoodeh, M.; Lavington, S. H; Standring, M. The semantic binary relationship model of information. In *Research and Development in Information Retrieval*, The British Computer Society Workshop Series, University Press, Cambridge, 1984.

The Semantic Binary Relationship Model (SBRM) is a first-order formalism which combines an organizationally simple basis (i.e., binary relationships) with the capabilities of semantic networks and logical integrity and deduction rules. The aim is to permit the efficient modeling of practical enterprises in a DBMS context, while accommodating the requirements of knowledge-based systems. The theoretical foundations of the SBRM are described, with particular attention to inheritance hierarchies and rule representation. The low-level unit of SBRM information is the triple. A 4Mbyte associatively-accessed triple store is being constructed, and will form the heart of a smart information machine based on the SBRM.

Beckley, D.A.; Evens, M.W.; Raman, V.K. Multikey retrieval from K-d trees and quad-trees. *SIGMOD Rec. (USA)*, 14(4):291-301, Proceedings of ACM-SIGMOD 1985 International Conference on Management of Data 2, Austin, December 1985.

Associative file structures are potentially valuable for many database and artificial intelligence applications, but very little information is available to database designers trying to choose an appropriate file structure for a particular problem. This paper describes an experiment comparing the retrieval performance of K-d trees, quad-trees and flat files, using means of time and I/O operations. Five types of queries are used: exact match, partial match, range search, nearest neighbor, and best match. The database used in this study is a static medical database of half a million characters with the patient information removed. Results suggest that there is no one best type of file structure for all types of associative queries; quad trees dominated with some query classes, K-d trees with others.

Bic, L. Dataflow architectures for knowledge representation systems. In *AFIPS Conference Proceedings: 1985 National Computer Conference*, 15-18 June, edited by A S Wojcik, AFIPS Press, Reston, VA, 1985:287-295.

Most AI systems today are written in languages based on the sequential von Neumann model of computation and thus are not well suited to parallel processing. This paper proposes to use logic programming as a bridge between the high-level operations performed by knowledge representation systems based on semantic nets and highly-parallel

computer architectures. It is shown how knowledge recorded in semantic nets, including the corresponding information retrieval operations, may be expressed in terms of logic programming. Such representation may be processed in a data-driven manner and thus permits computer architectures consisting of very large numbers of independent processing elements to be exploited.

Biswas, G; Subramanian, V; Bezdek, J. C. A knowledge based system approach to document retrieval. *Second Conference on Artificial Intelligence Applications: The Engineering*, 11-13 Dec. 1985.

A discussion is presented of the design and implementation of a prototype document retrieval system using a knowledge-based approach. Both the domain-specific knowledge base and the inferencing schemes are based on a fuzzy-set-theoretic framework. Concepts that describe domain topics and relationships between concepts are captured in a knowledge base. Documents are represented as concept-weight pairs and clustered into a few general subsets. The retrieval mechanism uses a two-step approach. In the first step, a pruned list of documents pertinent to the query is produced. In the second step, an evidence combination scheme is used to compute a degree of relevance between the query and individual document retrieved in step 1. Finally, a set of document citations are presented to the user in ranked order as an answer to his query.

Bose, P. K; Rajinikanth, M. KARMA: knowledge-based assistant to a database system. *Second Conference on Artificial Intelligence Applications: The Engineering of Knowledge-Based Systems*, Washington, DC, 11-13 Dec. 1985.

KARMA is a knowledge-based assistant to a relational database system. The system is designed to aid on the conceptual level, and guides toward the relevant information through a reformulation process. To assist users in retrieving the appropriate data, KARMA maintains a representation of the current partial description of the query and an individual example that satisfies the query. Information from these representations, as well as the knowledge base representing the concepts particular to the query domain, can be utilized to reformulate the query in an interactive fashion. The user-level query reformulation is viewed as analogous to the process of debugging a query specification at a conceptual level. By successful reformulations, the users, who may have a limited knowledge of the underlying database schema, can retrieve what they want. KARMA interprets the reformulated queries based on the knowledge about the database domain.

Brachman, R. J. A Structural Paradigm for Representing Knowledge. Able Publishing Corp, Norwood, NJ, 1984.

This book offers a unique perspective on almost a decade of research in the representation of knowledge in computational systems. In large part it duplicates the author's previous research, which presents an "associational network formalism for representing conceptual knowledge." However, each chapter is now accompanied by a retrospective section that follows the development of the initial ideas through their implementation, use, and clarification in subsequent years. This provides a unique opportunity for insight into better understanding of the semantics of a popular type of system for representing knowledge in a computer -- semantic networks. This led directly to the development of the KL-ONE knowledge representation system, a significant representation system that has influenced work in the field and has been found useful in numerous applications.

Brooks, H. M; Daniels, P. J.; Belkin, N. J. Problem descriptions and user models: developing an intelligent interface for dn *Advances in Intelligent Retrieval: INFORMATICS 8. Proceedings of an Aslib/British Computer Society Joint Conference*, 16-17 April, Oxford, England. Aslib: London, 191-214.

Any intelligent interface for document retrieval must simulate the functional performance of good human intermediaries. A number of human-human information interactions in the document retrieval situation are analyzed to specify two such model-building functions, user modelling and problem description, and to identify the kinds and extent of knowledge these functions need to perform successfully. Results indicate that both functions are necessary for successful intermediary performance, that they can be subdivided into discrete subfunctions which relate to one another primarily according to type of knowledge representation required and type of processing necessary, that knowledge sources for these functions are extensive, and that a distributed expert blackboard architecture will probably be required for realistic implementation of an intelligent interface.

Brownstein, M. Managing information intelligently (Quantum's Knowledge Management System). *Hardcopy*, 14(11):139-141, November 1985.

A new application of artificial intelligence technology has resulted in the development of a system that its developers report is capable of 'understanding' content and context of textual materials. This technology is what Quantum Development Corp. refers to as Knowledge Management. At first glance, the Knowledge Management System (KMS) appears to do the seemingly impossible (for a computer, that is) - it can store and understand vast amounts of material (reportedly more than 250 million documents), within each knowledge management topic. The system is said to be able

to quickly index and retrieve information from a large knowledge base, and can provide contextually appropriate answers to natural language queries.

Burger, R. H. Artificial intelligence and authority control. *Library Resources & Technical Services*, 28(4):337-345, October 1984.

Artificial intelligence (AI) is already part of the cataloging world. To support this contention, four AI concepts that have relevance for information retrieval systems are discussed and applied to the area of authority control in automated catalogs. Existing automated authority control systems are then analyzed, using two other AI concepts, augmentation and delegation. In conclusion, several implications of the relationship between AI and authority control are drawn.

Cavanagh, J. M. A. Intra-active retrieval systems. In *Learned Information*, National online meeting proceedings, 6-8 May 1986, Learned Information: Medford NJ, 59-65, 1986.

The underlying literature of artificial intelligence (AI) has a significant representation of work in the areas of judgement and decision-making. As an 'intuitive statistician', man has proved to be rather inept. Models of man, using fairly simple information-combining rules such as linear equations, have performed remarkable well in a variety of evaluative tasks. Such findings in the behavioral science have significance for the design of information retrieval systems. An attempt is made to review AI concepts as they bear upon information systems design. Particular emphasis is given to 'intra-active' systems, i.e. retrieval systems which are capable of generating or of modifying search formulations automatically.

Chignell, M. H.; Loewenthal, A; Hancock, P. A. Intelligent interface design. *IEEE 1985 Proceedings of the International Conference on Cybernetics and Soc*, 12-15 November 1985, 620-3.

An intelligent interface is defined as one where display and control functions are carried out by communication between a human. The design is illustrated using the example of online information retrieval. In this problem domain, the intelligent entity which communicates with the human (user) is an expert system search intermediary. Intelligent interfaces require extensive input from the fields of human factors and artificial intelligence.

Cooper, W. S. Bridging the gap between AI and IR (Artificial intelligence and information ret.) In *Research and Development in Information Retrieval*, The British Computer Society Workshop Series, Cambridge University Press, Cambridge, 1984.

Information retrieval, in the broadest sense of the term, includes a concern with 'expert' or 'knowledge-based' systems and their potential future successors. It is unlikely that sophisticated systems of this sort can be developed in such a way as to use an entire natural language without the assistance of an advanced, unified theory of language. The need for and probable character of such a theory are discussed.

Cortes, U.; Lopez de Mantaras, R; Manero, J; Plazy, E; Salra, Implementation of a system of diffuse search in a bibliographical database. *Consorci Inf. & Documentacio de Catalunya, Barcelona, Spain, 17-18 Oct. 1985.*

This communication describes the implementation of a search module for a bibliographical database designed around a standard format (MARC) and original software. The principle applied is the use of fuzzy instructions which interact with the data structure, which is generated according to the ALIS module. The paper describes the relationship structure of the data from the points of view of logic and physical arrangement, the structure of the user/machine interface, and finally the process of search.

Cross, G. R.; deBessonnet, C. G. Representation of Legal Knowledge for Conceptual Retrieval. *Information Processing and Management*, 21(1):35-44, 1985.

Describes traditional legal information retrieval systems--Juris, Lexis, Westlaw -- and several new rule-based, knowledge-based, legal knowledge reasoning, and analytical legal information systems -- Waterman and Peterson's Legal Decisionmaking System, Hafner's Legal Information Retrieval System, McCarty's TAMAN, and the deBessonnet representation of the Louisiana Civil Code.

deBessonnet, C. G.; Gross, G. R.; Hintze, S. J.; Rajinkanth, M.; Allen, R. F. Formalization of legal information. In *Proc. International Conference on Data Bases in the Humanities and Social Sciences 1983*, New Brunswick, New Jersey. Paradigm Press: Osprey FL, 425-434.

The research described in the paper is being conducted by the authors in the fields of scientific codification and artificial intelligence. The goal of the project is to establish an interactive retrieval system centered around the Civil Code of Louisiana. The provisions of the Civil Code pertaining to the law of property are serving as the initial domain of the project. Those rules were chosen because the property law of Louisiana has been recently revised. The revision has produced a set of rules that is more receptive to formalization than were the rules of the past because most of the imprecisions that existed under prior law have been removed in the process of revision.

Diener, R. A. V. Relational knowledge structures: a structural model of information for research and retrieval. In *Proc. Challenges to an Information Society*. Proceedings of the 47th ASIS Annual Meeting, Philadelphia, PA, October 21-25, 1984. Knowledge Industry Publications, Inc.: White Plains, NY.

The largely ignored practice of relational indexing is resurrected and refurbished as a Relational Knowledge Structure (RKS) model of information and knowledge. The model envisions information as being composed of concepts structured by relations. The Relational Knowledge Structure is being proposed as a method for developing knowledge bases for information retrieving systems and artificial intelligence applications; as well as the basis for the empirical measurement of information, an information metric (informetric) for a Science of Information.

Dryzek, H. Automatic indexing of documents in the MSIS NIR system. *Aktual. Probl. Inf. & Dok. (Poland)*, 30(6):19-23, Centrum INTE, Warszawa, Poland, 1985.

The method of automatic indexing of Russian-language documents in the frame of the polythematic database of the MSIS NIR system (Information Specialized Information System on Research Reports) is presented. The automatic indexing is carried out with the aid of a thesaurus and concerns the document title and abstract. The application of a foreign-language version of the basic thesaurus makes possible the obtention of the search pattern of a document in the given language.

Dubois, J E; Sobel, Y. DARC system for documentation and artificial intelligence in chemistry. *J. Chem. Inf. & Comput. Sci. (USA)*, 25(3):326-333, August 1985.

The DARC System deals with structural information both for documentation and for artificial intelligence (AI) endeavors in chemistry. Its topological concepts are briefly reviewed in conjunction with the creative data needs in knowledge information processing systems (KIPS). Knowledge base, inference engine, and user interface are discussed with reference to the DARC potential in the field of AI and expert systems. AI methodology and its impact on knowledge production are reviewed. New chemical computer-aided design (CAD) tools to develop more creative and innovative research in synthesis planning, structure elucidation, and prediction in drug design are no longer pure prospective challenges.

Estall, C. Shared processing with an advanced intelligent terminal. In *Research and Development in Information Retrieval*, The British Computer Society Workshop, edited by C J van Rijsbergen, Cambridge University Press, Cambridge, 1984.

The authors have built a prototype distributed information retrieval system known as TBIRD, based on an inverted file and shared between a personal computer, acting as an advanced intelligent terminal, and a timeshared mainframe. It was developed to study the response and cost in comparison with a conventional system based on an unintelligent terminal. It is shown, by the transfer of most of the processing to the personal computer, that the computing costs can be reduced by a substantial factor and that the response time need not be degraded except when the mainframe is lightly used or when the communications channel between the processors is slow (< 2,400 bps).

Fuhr, N. Retrieval test evaluation of a rule based automatic indexing (AIR/PHYS). In *Research and Development in Information Retrieval*, The British Computer Society Workshop, edited by C J van Rijsbergen, Cambridge University Press, Cambridge, 1984.

The automatic indexing system AIR/PHYS and its evaluation by means of a retrieval test with 309 requests and 15,000 documents is described. First, the underlying conception of a rule based approach is given which is suited to the task of a controlled-vocabulary indexing of even large subject fields. Preconditions, performance and results of the retrieval test are described, including first results of retrieval runs with weighted automatic indexing

Hahn, U. Expert systems as intelligent information systems. Perspectives for functional extension of information retrieval. *Nachrichten fuer Dokumentation*, 36(1):2-12, 1985.

A comparison of conventional document retrieval systems, fact retrieval, and expert systems is given on a functional level. Based on a legal information problem and illustrative interactions among system and user knowledge sources available for each system type as well as the background knowledge necessary on the users side are outlined. An illustration of expert systems' problem solving capabilities is provided applying deductive reasoning mechanisms to a legal rule base. The discussion of properties of conventional and intelligent information systems introduced so far leads to an estimation of how the development of expert information retrieval systems might influence current information practice.

Helbig, H. Natural language access to the data base of the AIDOS/VS information retrieval system. Elsevier Science Publishing Company, New York, 1984.

Natural language interfaces (NLI) mark one borderline between research in the field of Artificial Intelligence and its commercial application. Considering the fact that in the next decade the so-called "naive users" will become typical users of information systems, the realization of natural language access to data bases will become a necessity. This paper discusses the structure of a NLI for the AIDOS/VS information retrieval system.

Herther, N.K. PROLOG to the future: a glimpse of things to come in artificial intelligence. Microcomput. Inf. Manage. (USA), 3(1):31-45, March 1986.

Artificial Intelligence (AI), a science fiction theme for hundreds of years, has been the object of scientific research for over four decades. With the recent development of faster, more powerful computer systems, AI applications are beginning to appear in various products for information management and retrieval. This important new field has its own particular methodologies and programming languages; library/information professionals hoping to understand and use AI products will need to learn more about them. This article briefly introduces the languages of AI and presents information on some of the new versions of these languages available for personal computers. A tutorial for PROLOG-86, a new microcomputer version of PROLOG, is given. Information on other microcomputer versions of these programs and an extensive bibliography are included.

Jones, K. P.; Bell, C. L. M. MORPHS-an intelligent retrieval system. ASLIB Proc. (GB), 38(3):71-79, March 1986.

It may appear rash to claim that an information retrieval system is 'intelligent', even if such claims are made within the limited context of artificial intelligence; nevertheless, such a claim is implicit in the MORPHS acronym: minicomputer operated retrieval (partially heuristic) system. Heuristics are systems where the instigator is less than completely certain about the outcome of material fed into them.

Kleinbart, P. Prolegomena intelligent thesaurus software. Inf. Sci. Princ. & Pract., 11(2):45-53, 1985.

Since its appearance in the late fifties as an instrument for mechanized information retrieval, the thesaurus often has been subjected to reconsideration and scrutiny. The typical approach found in the literature on thesauri is to define a thesaurus historically, by giving its etymology. In order to go beyond this type of approach the thesaurus should be redefined by incorporating recent advances in such diverse fields as linguistics, artificial intelligence, programming techniques and compute architecture.

Klingbiel, P. H. Phrase structure rewrite systems in information retrieval. Information Processing and Management, 21(2):113-126, 1985.

Operational level automatic indexing requires an efficient means of normalizing natural language phrases. Subject switching requires an efficient means of translating one set of authorized terms to another. A phrase structure rewrite system called a Lexical Dictionary that performs these functions is explained. Background, operational use, other applications and ongoing research are explained.

Kristalnyi, B. V.; Voiskunskii, V. G.; Raskina, A. A.; Sidorov, I. S.; Sharygin, V. I. The 'intellectual capacity' of automated factographic information retrieval sys. Nauchno-Tekh. Inf. Ser., 2(12):4-11, 1985.

The place of problem-oriented automatic factographic information retrieval systems in the overall 'man-machine' concept is examined, and a particular application of the results of research into artificial intelligence systems to represent knowledge, and languages to represent data are discussed, and distinctions are drawn between pre- and post-retrieval intelligent information systems.

Kuhlen, R. Natural language research. SIGART Newsletter, mNo. 83: 20-21, January 1983.

TOPIC (Text-Oriented Procedures for Information Management and Condensation of Expository Texts) is an automatic text analysis system currently under development at the information science department of the University of Constance (West Germany). TOPIC will feature the condensation, i.e., abstracting of German language full texts (complete journal papers, scientific reports etc.) in the domain of in communication technology for business and office automation. TOPIC takes up the research tradition of "automatic abstracting" which stopped in the mid-sixties. Research in this area has meanwhile received impetus from two directions. On the one hand an enormous number of texts are generated in machine-readable form today (by editing and text systems, teletex, electronic mail etc.) and managed in large digital memories (text depots in terms of "publishing on demand"). On the other hand work in Artificial Intelligence has led to the development of quite sophisticated and powerful representation languages (semantic networks, frames, CD graphs etc.).

Kwok, K. L. A document-document similarity measure based on cited titles and probability theory, and its application to relevance feedback retrieval. In *Research and Development in Information Retrieval*, The British Computer Society Workshop Series, University Press, Cambridge, 1984.

The use of cited title terms of a scientific document for automatic indexing is explored. It offers a means of index term selection, as well as term relevance weighting, based on author-provided relevance information and Bayes Theorem as in probabilistic retrieval. The latter quantitative consideration leads to a new measure of document-document similarity measure which is shown to have importance both for initial search and in relevance feedback retrieval, by offering a choice of iterative strategies.

Leigh, W; Evans, J. Interpretation of natural language database queries using optimization methods. *IEEE Trans. Syst., Man & Cybern. (USA)*, SMC-16(1):40-52, Jan.-Feb. 1986.

The automatic interpretation of natural language (English) database questions formulated by a user untrained in the technical aspects of database querying is an established problem in artificial intelligence. State-of-the-art approaches involve the analysis of queries with syntactic and semantic grammars expressed in phrase structure grammar or transition network formalisms. With such methods difficulties exist with the detection and resolution of ambiguity, with the misinterpretation possibilities inherent in finite length look-ahead, and with the modification and extension of a mechanism for other sources of semantic knowledge. The potential of optimization techniques to solve these problems and interpret natural language database queries is examined.

McCune, B. P.; Tong, R. M.; Dean, J. S.; Shapiro, D. G. RUBRIC: a system for rule-based information retrieval. *IEEE Trans. Software Eng.*, SE-11(9):939-945, September 1985.

A research prototype software system for conceptual information retrieval has been developed. The goal of the system, called RUBRIC, is to provide more automated and relevant access to unformatted textual databases. The approach is to use production rules from artificial intelligence to define a hierarchy of retrieval subtopics, with fuzzy content expressions and specific word phrases at the bottom. RUBRIC allows the definition of detailed queries starting at a conceptual level, partial matching of a query and a document, selection of only the highest ranked documents for presentation to the user, and detailed explanation of how and why a particular document was selected. Initial experiments indicate that a RUBRIC rule set better matches human retrieval judgement than a standard Boolean keyword expression, given equal amounts of effort in defining each.

McGregor, D. R.; Malone, J. R. An architectural approach to advances in information retrieval. In *Advances in Intelligent Retrieval: INFORMATICS 8. Proceedings of an Aslib/British Computer Society Joint Conference, 16-17 April, Aslib, London, 34-46, 1985.*

The Fact system is based on a simple representation of knowledge, the elements of which are represented as individual molecules. Each 'molecule' is a four-tuple, three fields of which represent a link in a semantic network, and the fourth is a 'name' or 'label' which can represent the unit of information itself in other 'molecules'. (The system is thus capable of representing higher-order logics.) The Generic Associative Memory (GAM) device can be applied to any information handling system. In the context of the Fact system it can evaluate the complete closure of 'semi-explicit' (finite) sets; it can store any desired network, which can be set-up dynamically as a transitive, directed graph; it can rapidly locate the physical storage blocks in which particular elements of data have been stored; and it can be used for high-speed systems such as real-time vision understanding.

Medler, N. Artificial intelligence as a tool of classification, or: the network of language. *Int. Classif. (Germany)*, 12(3):128-132, 1985.

It is shown that the cognitive paradigm may be an orientation mark for automatic classification. On the basis of research in artificial intelligence, the cognitive paradigm - as opposed to the behavioristic paradigm - was developed as a multiplicity of competitive world-views. Multiplicity in a loosely-coupled network of cognitive knots is also the principle of dynamic restlessness. In competition with cognitive views, a classification system that follows various models may learn by concrete information retrieval. During his actions the user implicitly builds a new classification order.

Merry, M. *Expert Systems 85. Proceedings of the Fifth Technical Conference of the British Computer Society Specialist Group on Expert Systems. Cambridge Univ. Press, Cambridge, England, 1985.*

The following topics were dealt with: problems and opportunities with expert systems; diagnostic expert systems; inference engine choice; expert-system question selection; a statistics application; a hard-wired circuit troubleshooting application; ESCORT, a process control system; knowledge-based planning techniques; multiple-agent computer system planning; choice-making in planning; ECO, a browsing system; CAD/CAPP (CAD and computer-aided process planning); and symbolic uncertain inference.

Mihram, G. A.; Mihram, D. Artificial intelligence and the fifth generation: some wrong ways, the correct way. In *Proceedings of the fourteenth annual computer science conference, RCSC 1986*, 4-6 February 1986, ACM, New York, 1986.

The widespread acclaim for artificial intelligence (AI) research in academia is underscored by the realisation not only that every computer program is in actuality a simulation model, but also that the automated data retrieval 'languages' being used within and among libraries are actually AI models (i.e. simulations or algorithmic models) of the mind of a librarian researcher engaged in the pursuit of published information on a particular topic at hand. AI is the effort to prepare and to present scientifically credible models of the mind, typically of the adult and human mind, at work at one or more of the tasks before it.

Mitev, N. N.; Walker, S. Information retrieval aids in an online public access catalogue: automatic inte. In *Advances in Intelligent Retrieval: INFORMATICS 8. Proceedings of an Aslib/British Computer Society Joint Conference, 16-17 April 1985 Oxford, England, Aslib, London, 215-26, 1985.*

In conventional IR systems the burden is on user/intermediary to reformulate queries until satisfactory results are obtained. Using an online public access catalogue (OPAC), attempts have been made to devise rules for using the result of a search so far to determine its course. These sequencing rules take account of the degree of similarity of items in the query to index terms, and their frequency. Tentative steps have been made towards formulating some general principles.

Morrissey, J. M. Interactive Querying Techniques for an Office Filing Facility. *Information Processing and Management*, 22(2):121-34, 1986.

Proposes a "Model of Querying" for users of office filing facilities and discusses its motivation, aspects, attributes, and advantages. A review of current information systems and attempts to combine information retrieval, artificial intelligence, and database management techniques leads to conclusion that no resultant system is adequate for office environments.

Nowak, E. J.; Szablowski, B. F. Expert Systems in Scientific Information Exchange. *Journal of Information Science: Principles and Practices*, 8(3):103-111, 1984.

The paper presents expert systems as a new generation of information storage and retrieval systems which may considerably improve the processes of scientific and technical information exchange and dissemination. Essential features of the expert systems with databases containing scientific or technical information have been specified. For databases in which a semantic network is used as a knowledge representation scheme, an idea of database organization has been described with some attention paid to the problem of extracting factual information from the texts of scientific and technical publications. The process of information retrieval from the database of a scientific or technical information expert system is also briefly described.

Rada, R.; Humphrey, S.; Suh, M.; Brown, E.; Coccia, C. Relevance on a biomedical classification structure In *Proc. Expert Systems in Government Symposium*, 24-25 Oct. 1985 McLean, VA, USA, IEEE Comput Soc. Press, 1985.

The authors have developed algorithms that reference a semantic network and calculate the conceptual closeness between a document and a query. The algorithms were used to calculate closeness between MeSH-encoded documents and the MeSH search words taken for the query. It was found that significant agreement occurred between the computer evaluations and those by the querist. The authors sketch some of the mathematical and data-retrieval properties of one of their computer algorithms. The mathematical properties help one anticipate the effect of changes to query and document classifications in terms of retrieval. The data-retrieval properties are also presented.

Restorick, Mark F. Novel Filing Systems Applicable to an Automated Office: A State-of-the-Art Study. *Information Processing and Management*, 22(2):151-72, 1986.

Examines novel computer filing systems which have particular application to office information storage and retrieval requirements. A variety of filing systems and their major characteristics are reviewed, ranging from network-based file servers to digital image storage and retrieval systems. Desirable characteristics of a modern electronic office filing system are discussed. (Author/MBR)

Rumelhart, D. E. Representation in Memory. UCSD No ED 235-770 Thesis, UCSD, June 1983.

This paper reviews work on the representation of knowledge from within psychology and artificial intelligence. The work covers the nature of representation, the distinction between the represented world and the representing world, and significant issues concerned with propositional, analogical, and superpositional representations. Specific topics related to propositionally-based systems are semantic features or attributes, symbolic logic and the predicate calculus, semantic networks and their properties, schemata and frames, and the relationship of these representations to classical associations. A section on analogical representations discusses the work of Shepard, Kosslyn, and Funt; mental models

and mental simulations; and propositional and analogical representation. Procedural representation topics include the human information processing system, procedural representation, and demons and production systems. Additional topics include local and superpositional memory systems, associative memories, cognition and categorization, and generalization. Major controversies within psychology -- such as distinctions between declarative and procedural representation, propositional and analogical representation and the nature of visual images -- are analyzed. (Abstract Source: ERIC)

Salton, G. Some characteristics of future information systems. *SIGIR Forum (USA)*, 18(2-4):28-39, Fall 1985.

Existing information systems are characterized by sophisticated hardware designs and relatively unforgiving software support. One may expect that future information systems will provide a unified approach to several different types of information processing tasks, as well as more user-friendly processing environments. Some of the requirements of future information systems are described and various advances in retrieval system design are examined, including automatic indexing, automatic query formulation, and extended Boolean query processing.

Saparty, P. S. Solving tasks on semantic networks and graphs by active distributed structures. In *Artificial Intelligence and Information Control Systems for Robots*, Elsevier Science Publishing Co, New York, 1984.

Efficient methods of structural solving of complex tasks on semantic networks and graphs are presented. They are based on a concept of active data networks without separating data representation and computation. The data networks find solutions in a self-organizing manner with high degree of performance when put into a multiprocessor environment. A hypothetical network computer is discussed for the direct interpretation of an active data network model.

Sellis, T. K.; Shapiro, L. Optimization of extended database query languages. *SIGMOD Rec. (USA)*, 14(4):424-436, Proceedings of ACM-SIGMOD 1985 International Conference on Management of Data 28-31 May 1985, Austin, December 1985.

The authors examine the problem of query optimization for extended data manipulation languages. They propose a set of tactics that can be used in optimizing sequences of data base operations and describe the corresponding transformation procedures. These transformations result in new equivalent sequences with better space and time performance. The proposed techniques are especially useful in artificial intelligence and engineering applications where sequences of commands are executed over high volume databases.

Shaw, W. M., Jr. An Investigation of Document Partitions. *Information Processing and Management*, 22 (1):19-28, 1986.

Empirical significance of document partitions is investigated as a function of index term-weight and similarity thresholds. Results show the same empirically preferred partitions can be detected by two independent strategies: an analysis of cluster-based retrieval analysis and an analysis of regularities in the underlying structure of the document graph. (Author)

Shoval, P. Principles, procedures and rules in an expert system for information retrieval. *Inf. Process. & Manage*, 21(6):475-487, 1985.

An expert system was developed in the area of information retrieval, with the objective of performing the job of an information specialist, who assists users in selecting the right vocabulary terms for a database search. The system is composed of two components: the knowledge base, represented as a semantic network, in which the nodes and words, concepts, and phrases, comprising a vocabulary of the application area and the links express semantic relationships between those nodes; and rules or procedures, which operate upon the knowledge-base, analogous to the decision rules or work patterns of the information specialist. Two major stages comprise the consulting process of the system: during the search stage relevant knowledge in the semantic network is activated, and search and evaluation rules are applied in order to find appropriate vocabulary terms to represent the user's problem; during the suggest stage those terms are further evaluated, dynamically rank-ordered according to relevancy, and suggested to the user. Explanations to the findings can be provided by the system and backtracking is possible in order to find alternatives in case some suggested term is rejected by the user. This article presents the principle, procedures and rules which are utilized in the expert system.

Smeaton, A. F.; van Rijsbergen, C. J. Information Retrieval in an Office Filing Facility and Future Work in Project Minstrel. *Information Processing and Management*, 22(2):135-49, 1986.

Review of office filing facility filing and retrieval mechanisms for unstructured and mixed media information focuses on free text methods. Also discussed are the state of the art in handling voice and image data, problems with searching text surrogates to implement free text content retrieval, and work of Project Minstrel. (Author/MBR)

Veith, R. H. Information retrieval and spatial orientation. In *ASIS '85. Proceedings of the 48th ASIS Annual Meeting, 20-24 Oct. 1985, Las Vegas, NV, 250-254, 1985*

Prior to the advent of computerized databases, the location and arrangement of information often had a very real spatial orientation - information was located on that piece of paper, on this shelf, in that room. Given the continuing developments in processing and display technologies, it may be time to once again provide spatial orientation as a fundamental feature of retrieval systems. This paper argues that such systems would have improved usability while at the same time building on much of the work done on associative procedures, and even artificial intelligence or expert systems, for retrieval purposes. In short, now that some systems have icons and windows, it is time to add maps.

Vickery, A. An intelligent interface for online interaction. *Journal of Information Science: Principles & Practices*, 9(1):7-18, 7-18 August 1984.

In this paper, the author discusses the ways of improving the performance of online retrieval systems by introducing an automated interface between the enquirer and the system. In the first part of the paper, the main features of such human/machine interaction and the characteristics that the user would like to see incorporated in an interface, are described. Then, studies in artificial intelligence that are particularly relevant to the problems of implementing an intelligent interface, are discussed. The author concludes with a summary of automated mechanisms that will be needed to improve the quality of interaction between the user and the search system.

Walker, D. E. Knowledge resource tools for accessing large text files: Artificial Intelligence & Inf. Sci. Res. (Bell Commun. Res). In *Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, 14-16 Aug. 1985* Hamilton, NY, USA, edited by S. Nirenburg, Colgate Univ, Hamilton, NY, 335-347, 1985.

This paper provides an overview of a research program defined at Bellcore. The objective is to develop facilities for working with large document collections that provide more refined access to the information contained in these 'source' materials than is possible through current information retrieval procedures. The tools being used for this purpose are machine-readable dictionaries, encyclopedias, and related 'resources' that provide geographical, biog knowledge. A major feature of the research program is the exploitation of the reciprocal relationships between sources and resources. These interactions between texts and tools are intended to support experts who organize and use information in a workstation environment. Two systems under development are described to illustrate the approach: one providing capabilities for full-text subject assessment; the other for concept elaboration while reading text. Progress in the research depends critically on developments in artificial intelligence, computational linguistics, and information science to provide a scientific base, and on software engineering, database management, and distributed systems to provide the technology.

Wilson, J. H. Retrieving, structuring, and presenting knowledge from very large bibliographical databases. In *Proceedings of the Eighteenth Hawaii International Conference on System Sciences 1985, 2-4 Jan. 1985* Honolulu, HI, 2:798-801

The orientation of bibliographical systems is verbal - texts are in words. But word relationships include more than is treated in the traditional treatment of continuous text. In approaching large operating storage and retrieval systems, inputting by computer and having on-line access, with a view to structuring and presenting knowledge other 'grammatical' elements have to be considered. The knowledge to be retrieved, structured, a presented is not just in the words of the text. The overall problem of structuring and presenting knowledge from storage and retrieval systems is a variety of the effort in artificial intelligence to backtrack from the goal state to the initial state. However, with large bibliographical systems the problem is somewhat different: specifically to retrieve and then structure and present what was not necessarily a specific, isolated intent or content of the documents or their surrogates when entered into the system. Even if the information is there, the knowledge is projected as never conceived by the authors of the documents.

Winnett, S. G.; Fox, E. A. Using information retrieval techniques in an expert system. In *Second Conference on Artificial Intelligence Applications: The Engineering of Knowledge-Based Systems*, 11-13 Dec. 1985 Miami Beach, FL, IEEE Comput. Soc. Press, 230-235, 1985.

The Foster Care Expert System (FOCES) is being developed to provide advice to social workers at the Roanoke, Virginia City Department of Social Services. Its primary aim is to find appropriate matches between a child and possible foster care sites. FOCES integrates advanced information retrieval (IR) techniques with a PROLOG-based General Purpose Expert System Shell (GUESS). Evaluation of the effectiveness of this approach should lead to a better understanding of the utility of IR methods in expert systems.

Wong, S. K. M.; Ziarko, W. On generalized vector space model in information retrieval. *Ann. Soc. Math. Pol. Ser. IV: Fundam. Inf. (Poland)*, 8(2):253-267, 1985.

In information retrieval, it is common to model index terms and documents as vectors in a suitably defined vector space. The main difficulty with this approach is that the explicit representation of term vectors is not known a priori. For this reason, the vector space model adopted by G. Salton for the SMART system (The SMART Retrieval System-Experiments in Automatic Document Processing, Prentice-Hall, Englewood Cliffs, New Jersey, 1971) treats the terms as a set of orthogonal vectors. But is often necessary to adopt a separate, corrective procedure to take into account the correlations between terms. This paper proposes a systematic method (the generalized vector space model) to compute term correlations directly from the automatic indexing scheme. It also demonstrates how such correlations can be included with minimal modification in existing vector based information retrieval systems.

Zarri, G. P. Interactive information retrieval: an artificial intelligence approach to deal. In *Advances in Intelligent Retrieval: INFORMATICS 8. Proceedings of an Aslib/British Computer Society Joint Conference*, 16-17 April 1985 Oxford, England, Aslib, London, 101-19, 1985.

Provides some information on RESEDA, an 'intelligent' information retrieval system working on factual databases containing biographical data. This data is described using a particular knowledge representation language ('meta language') based on an artificial intelligence understanding of a case grammar approach. The computing kernel of the system consists of an inference engine, which uses two types of rules: transformations and hypotheses. Via the hypotheses, the system is able to automatically establish new causal links between the events represented in the factual database.

Zarri, G. P. Some remarks about the inference techniques of RESEDA, an "intelligent" information retrieval system. In *Research and Development in Information Retrieval*, The British Computer Society Workshop Series, University Press, Cambridge, 1984.

The aim of this paper is to provide some details about the inference procedures of RESEDA, an "intelligent" Information Retrieval system using techniques borrowed from Knowledge Engineering. A RESEDA prototype has become operational: among its characteristics are a "case grammar" like knowledge representation language, intensive use of temporal data, use of the notion of "type," automatic generation makes it clear what is meant in RESEDA by "level zero influence." Subsequently, it provides an informal description of the two kinds of high level inference operation, relying on information in the rule base, that are implemented in the system: these are known as "transformations" and "hypotheses." Finally, the article describes in some detail the computational structure of the "machines" which enable RESEDA's inference engine to execute this type of high level operation.