Doszkocs, T., & Sass, R. (1992). An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching. 3rd ASIS SIG/CR Classification Research Workshop, 12-36. doi:10.7152/acro.v3i1.12593

PROCEEDINGS OF THE 3rd ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching

Tamas E. Doszkocs
National Library of Medicine
Office of Computer and Communications Systems
8600 Rockville Pike, Bethesda, Md. 20894

Rivkah K. Sass
Washington State Library
Library Planning and Development Division
P.O.Box 42472, Olympia, WA 98504-2472

"In the beginning was the word"

Capturing and exploiting textual database associations has played a pivotal role in the evolution of automated information systems. A variety of statistical, linguistic and artificial intelligence approaches have been described in the literature.

Many of these R & D concepts and techniques are now being incorporated into commercially available search systems and services.

This paper discusses prior work and reports on research in progress aimed at creating and utilizing a global semantic associative database, **AURA (Associative User Retrieval Aid)** to facilitate machine-assisted indexing, classification and searching in the large-scale information processing environment of NLM's core bibliographic databases, MEDLINE and CATLINE.

AURA is a semantic network of over two million natural language phrases derived from more than a million MEDLINE titles. These natural language phrases are associatively linked to NLM's MeSH (Medical Subject Headings) and UMLS Metathesaurus (Unified Medical Language System) controlled vocabulary and classification resources.

1. INTRODUCTION

Research and development in information retrieval over the past three decades has been characterized by the predominant use of statistical and probabilistic approaches, and to a lesser extent, linguistic and artificial intelligence methods and techniques, in all major aspects of the retrieval process, namely query and document analysis, document classification, query-to-document matching and query and database modification and restructuring [Sparck-Jones (1973), Salton (1983), Salton (1989), Larson (1991)]. In essence, **these approaches involve associative information processing** via the identification and explicit use of term-term associations (automatic thesaurus construction), document-document associations (automatic classification), term-document associations (automatic indexing) and query-document associations (automatic retrieval).

From the user's point of view, such systems offer considerable flexibility, including natural language query input, ranked display of closest matching items, and automatic query refinement and search strategy modification by the system in response to user feedback concerning the relevance of retrieved items.

ISSN: 2324-9773

Doszkocs, T., & Sass, R. (1992). An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching. 3rd ASIS SIG/CR Classification Research Workshop, 12-36. doi:10.7152/acro.v3i1.12593

By contrast, commercial retrieval systems and services have, until recently, limited themselves to rudimentary automatic keyword indexing (often supplemented by labor intensive human controlled vocabulary indexing), manual construction of machine-readable thesauri and classifications and the manual assignment of such thesaurus and controlled vocabulary terms to records to describe document content. The explicit matching of search queries (formulated as Boolean expressions) to documents and/or document surrogates is typically accomplished via the exact match processing of inverted lists corresponding to the search keys and the matching logic specified in the query. **These systems do not perform any automatic associative information processing**, beyond the passive use of the inherently associative aspects of the human intellectual effort involved in indexing and classification, and the explicit network of associations between cited and citing references, footnotes and similar traditional referential devices.

2. PRIOR WORK AND RECENT DEVELOPMENTS

2.1. Associative Information Processing

The human record is a reflection of mankind's collective memory and as such, it is a mirror of reality associations.

From the earliest days of computing, visionaries, self-avowed computopians and IR pioneers have explored a wide variety of associative approaches to information processing [Bush (1945)], , Doyle (1961), Stevens (1965), Stiles (1961), Lesk (1969), Nelson (1970), Sparck Jones (1971), Salton (1975), van Rijsbergen (1977), Doszkocs (1978)].

At their very best, information retrieval systems are hit-and-miss due to the fundamental mismatch between the language of authors, indexers-catalogers and searchers, the wonderful richness, versatility and ambiguity of natural language and the consistent inconsistency of people in their writing, indexing, classification and searching [Doszkocs (1983), Borgman (1989), Belkin (1987), Egan (1988), Marchionini (1989), Allen (1991), Cleverdon (1991), Saracevic (1991), Vizine-Goetz (1991)]. In looking at the ample evidence provided by the literature in cognitive psychology, ergonomics and information retrieval concerning the fundamental limitations of both short-term and long-term memory in humans, and their limitless capacity for serendipity, imagination and creativity, one can reasonably expect that by providing associative displays of search terms and related documents warranted by the global database, an associative information retrieval system can ease the perennial problems of no retrieval, too little retrieval or too much retrieval from the user's subjective relevancy point of view.

The limitations of associative term co-occurrence data for automatic query expansion in document retrieval systems [Salton (1986), Peat (1991)] suggest that machine-assisted methods might be more appropriate and productive, as evidenced by research on associative interactive dictionary displays for user selection and human controlled query modification in a very large scale operational retrieval environment [Doszkocs (1979), Doszkocs (1983)]. Such findings are none too surprising, given that humans remain the most intelligent components of our systems, including our expert systems for retrieval, indexing and classification [Alberico (1989), Bates (1990), Belkin (1990), Brooks (1987), Croft (1987), Deerwester (1990), Fox (1987), Gauch (1989), Gibb (1988), Hawkins (1988), Hjerppe (1989), Humphrey (1989), Krulee (1989), Marcus

ISSN: 2324-9773

Doszkocs, T., & Sass, R. (1992). An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching. 3rd ASIS SIG/CR Classification Research Workshop, 12-36. doi:10.7152/acro.v3i1.12593

(1991) Meadows (1982), Micco (1987), Milstead (1992), Pejtersen (1987), Pejtersen (1989), Pollitt (1987), Richardson (1989), Roysdon (1989), Sanchez (1987), Sharif (1988), Smith (1987), Sparck Jones (1988), Vickery (1987)].

Associative information processing techniques have become an integral part of new and complementary information technologies, such as hypertext and hypermedia, a variety of artificial intelligence applications (e.g. expert systems and connectionist models) and scientific information visualization.

2.2. Hypertext and Hypermedia Retrieval

Much as indexing and document clustering result in an implicit network of related records and inherently facilitate nonlinear retrieval, hypertext and hypermedia [Nelson (1981), Nielsen (1990), Chen (1991)] support a network of explicit linkages among information objects and intrinsically facilitate **nonlinear associative browsing** [Egan (1989), Agosti (1991), Lesk(1991)]. The explicit linkages in hypertext and hypermedia systems are typically created via human intellectual assignment, be it the classical cited and citing references in bibliographic records or logical pointers to images, data tables or text elsewhere in the database. The automatic generation of "hyper-paths" in databases using a neural network algorithm has been proposed [Lelu (1991)].

2.3. Artificial Intelligence

AI knowledge representation and knowledge processing has had a profound influence on the development of "intelligent information retrieval systems" [Croft (1987), Fox (1987), Smith (1987), Turtle (1990)]. It is important to recognize that the (by and large handcrafted) **AI knowledge structures in essence represent rich associative relationships among the data, and the various AI processing methods, whether rule-based and procedural in nature, or ad hoc and intuitive, have a fundamental associative characteristic** [Partridge (1990)]. This observation also applies to practical natural language processing [Doszkocs(1986)], and particularly to the variety of applied "linguistic engineering" tools and resources, such as spellcheckers, computer dictionaries and thesauri, that are increasingly incorporated as integral components of information retrieval systems.

2.4. Connectionist Models

Connectionist models (artificial neural networks, spreading activation models, associative maps, parallel distributed processing) represent information as a network of weighted, interconnected nodes. In contrast to more conventional information processing methods, connectionist systems are "self processing", in that no external algorithm operates on the data in the network. The network literally processes itself, with "intelligent behavior" emerging from the local interactions that occur concurrently between the numerous network components [Reggia (1988)]. **Such models can explicitly and dynamically capture the rich variety of implicit associations in a database.** Connectionist models in information retrieval were surveyed by [Doszkocs (1990)]. Due to their inherent abstraction, generalization and adaptive learning capabilities, such systems promise to be very useful in emulating intuitive, human-like associative information processing. Nash (1989) implemented an expert system for thesaurus management using neural network principles,

ISSN: 2324-9773

Biennier (1990) employed a connectionist method to retrieve information in hyperdocuments, Sutcliffe (1991) described the use of vector space models in conjunction with connectionist ideas and natural language understanding, Kwok (1991) reported on associative query modification using an adaptive artificial neural network architecture, Wilkinson (1991) explored the use of the cosine similarity measure in a document retrieval system based on a neural network model, Lin (1991) implemented a self-organizing semantic association map using Kohonen's feature map, Waltz (1991) discussed parallel distributed processing R & D at Thinking Machines Corporation, Doszkocs (1991) assessed the potential of neural networks in libraries, Okaya (1991) presented the use of a neural network for terminology associations, and Lewis (1991) summarized research on adaptive intelligent information retrieval.

2.5. User Interfaces and Information Visualization

In topical searching, natural language is the least and the most common denominator among users and, in a very real sense, a few words are worth a thousand pictures in terms of user query input. At the same time, a picture is worth a thousand words on the system output side, when it comes to conveying information to the user about associative database content and relationships, and pathways to follow in vast and growing information spaces . Information retrieval systems today offer a dazzling variety of user interfaces, including natural language, command language, direct manipulation WIMP (windows, icons, menus, pointers), graphical, visual and exotic interfaces to suit a wide variety of problem and task domains, different types of databases and idiosyncratic users. Hybrid approaches include natural language query input, graphical user interfaces to restricted Boolean query formulation based on generalization/ aggregation hierarchies, filter/flow metaphors for complete Boolean expressions, dynamic query methods with continual visual representation and integrated navigational visual network structures for browsing associative query, document content and term relations [Fowler (1991), Henry (1991), Newby (1991), Schneiderman (1991)].

2.6. The State-of-the-Art in Information Retrieval Systems

During the 1960's and 70's, commercial search systems essentially ignored the achievements and recommendations of the IR research community. The usual reasons given were the small scale and lack of generalizability of experimental R & D systems. By the late 1980's, however, many of the established operational systems began to include capabilities pioneered by the research community.

We would like to illustrate the gradual incorporation of IR research into pragmatic applied development by examples from the National Library of Medicine:

The first large scale implementation of dynamic online associative index term displays, the **Associative Interactive Dictionary**, was implemented for the MEDLINE, TOXLINE and CHEMLINE databases at NLM in 1978 [Doszkocs (1978)].

The following year saw the first operational implementation of **CITE (Computerized Information Transfer in English)**, an inverted file based system with natural language query input, lexical-morphological and partial syntactic query analysis, limited AI knowledge structures

ISSN: 2324-9773

and retrieval heuristics, closest match, weighted logic search strategy, ranked output and dynamic associative query and search strategy modification based on user feedback and relevance judgements in searching the MEDLINE medical literature database, [Doszkocs (1979)]. It is important to point out that CITE used the same pre-existing inverted file data structures utilized by ELHILL, NLM's conventional, inverted file, Boolean search system.

A flexible **CITE Online Public Access Catalog** was made available for public use in the NLM Reading Room in 1983 [Doszkocs (1983)].

IRX, a state-of-the-art information retrieval system for experimentation was subsequently designed and implemented by another group of researchers at NLM [Harman (1988)].

The **UMLS** (Unified Medical Language System) Knowledge Sources, Experimental Edition, on CD-ROM, with Metathesaurus and Semantic network was released by NLM in September 1990 [Lindberg (1990)]. The **Metathesaurus (Meta-1)** contains 28,816 core concepts and related concepts, and 35,307 supplementary chemicals, for a total of 64,123 entries. Domain experts reviewed and enhanced the core and related concepts. The Meta-1 entries contain 208,559 lexically unique terms that include approximately 16,000 Medical Subject Headings from 1990 MeSH and also 1990 MeSH Supplementary Chemicals and terms from other major biomedical vocabulary sources. Organized as a relational database (and also as a hypercard stack for the Macintosh version), Meta-1 contains fields for the main concept and related information, synonyms, lexical variants, syntactic categories, semantic types and a variety of other data elements. Importantly, **Meta-1 includes approximately 4.3 million unique pairs of medical subject heading cooccurrences from approximately 3 million MEDLINE journal citations.** These co-occurrence records represent MeSH heading pairs that were assigned to the same citation. As such, these subject heading cooccurrences form a static network of topical associations that can potentially serve as an invaluable semantic user aid in formulating search strategies and either broadening or narrowing unsatisfactory searches.

The National Center for Biotechnology Information (NCBI) at the NLM released its experimental **Entrez: Sequences** CD-ROM database in 1991 [NCBI (1991)]. Entrez: Sequences is a retrieval system developed at the NCBI which provides an integrated approach for gaining access to nucleotide and protein sequence databases (Genbank, PIR and GenInfo) and to the approximately 86,000 MEDLINE citations in which the sequences were published. **A key feature of the system is the concept of 'neighboring', which permits a user to locate related references or sequences by asking Entrez to 'Find all papers that are like this one' or 'Find all sequences that are like this sequence'.** In effect, for each of these records a list of its nearest neighbors has been computed by treating the record as a query against the database using a cosine coefficient vector retrieval method [NCBI (1991)]. The relevant vectors are based on key terms coming from the titles, abstracts and MeSH headings associated with documents and weighted in a manner to take advantage of the relative importance of the different kinds of terms used. The top twenty documents retrieved become the neighborhood list for that document, and this information is also stored on the CD-ROM. While not all neighbors of a document found this way are guaranteed to be relevant, in initial testing, the first neighbor found is relevant to the query about 80% of the time. Fig. 1 shows an example of the nearest neighbors list from Entrez MEDLINE.

ISSN: 2324-9773

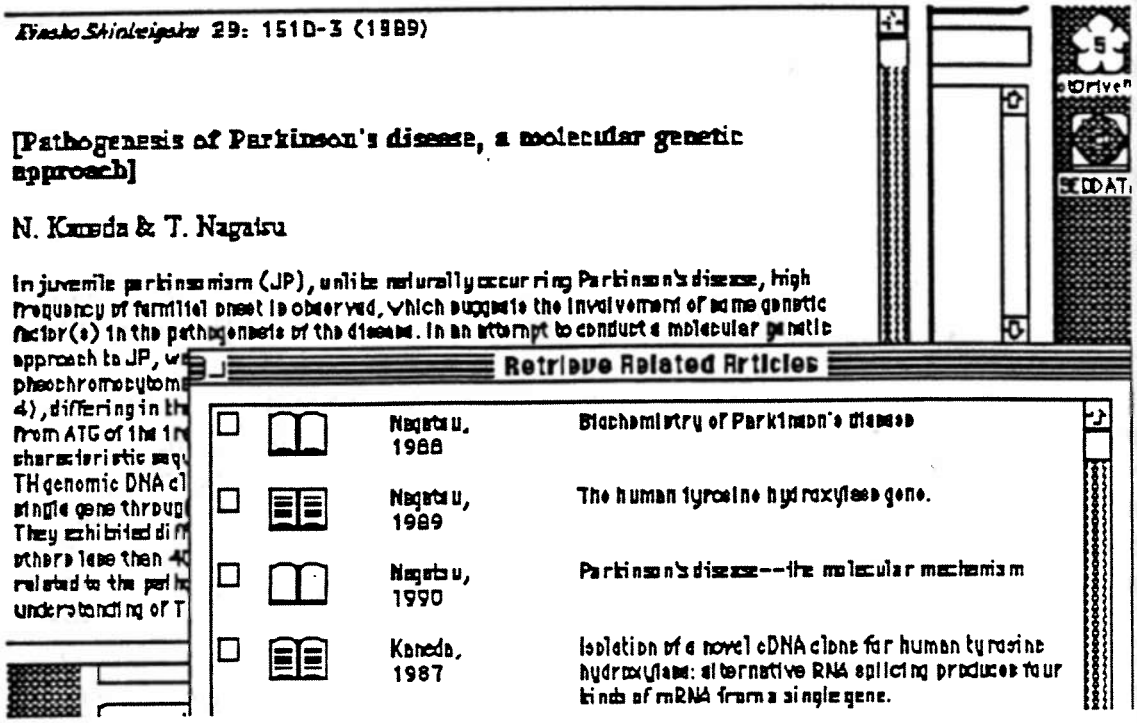


Figure 1. NLM NCBI MEDLINE Hypertext Links Among Preclustered Closest Matching Items

By the late 1980's, the unprecedented explosion in the mass utilization of personal computers and workstations, office automation, word processing, desktop and CD-ROM publishing, database management, E-mail, local area and wide area networking and public and personal information utilities have resulted in information retrieval becoming one of the "hottest" areas of software development and one of the best selling product categories in the marketplace.

It is gratifying to see that the latest generation of information retrieval systems have overcome most of the problem-of-scale limitations of early experimental IR, and representative systems, such as

DOW-QUEST (on the CONNECTION MACHINE)
ConQuest,
EXCALIBUR/SAVVY,
GESCAN,
INTELEQ/INFOTRACT,
KNOWLEDGE FINDER,
LMDS/LOGICON,
PL/PERSONAL LIBRARIAN,
SEARCH EXPRESS,
TRW/FAST DATA FINDER,
VERITY/TOPIC,
WAIS (The Wide Area Information Server of Thinking Machines Corporation)

ISSN: 2324-9773

and many others, incorporate most of the desirable features long ago advanced and championed by the IR research community [Levine (1991)]. Fig. 2 summarizes the features of the most advanced operational systems:

- **Hybrid Interfaces (NL, WIMP, GUI)**
  (natural language,windows-icons-menus-
- **pointers, graphical user interfaces)**
- **Linguistic Engineering**
- **Knowledge Structures**
- **Boolean, Fuzzy and Pattern Logic**
- **Ranked Output in Context**
- **Hypertext and Hypermedia**
- **Associative Query Modification**
- **Support for Standards**
- **Client Server Architecture**
- **Parallel Distributed Processing**
  " Non von Neumann ", " Connectionist "
  " Neural Network" search engines

Figure 2. Features of the Most Advanced IR Systems

3. A PRAGMATIC ASSOCIATIVE RETRIEVAL SYSTEM

During the current year, we initiated the prototype development of a large-scale, unified machine-assisted indexing, classification and retrieval system at NLM, based on the following assumptions:

a. large textual databases can be regarded as inherently associative knowledge bases.

b. the implicit associations in such databases can be made explicit by the use of well established automatic text processing methods, including pragmatic linguistic, statistical-probabilistic, knowledge-based, heuristic and neural network indexing approaches as applied to individual records, as well as globally to subsets of the database and the entire database itself [Humphrey (1989), Croft (1991), Fuhr (1991), Hersh (1991), Rau (1991), Warner (1991)]. Commercial products, such as AIDA, are now available that combine statistical, linguistic and heuristic techniques to perform intelligent text analysis and indexing [Jones (1991)].

c. Where available, controlled vocabularies (e.g. MeSH), classification systems (e.g. NLM Classification) and semantic knowledge bases (e.g. NLM UMLS METATHESAURUS), and their associative linkages with the subject databases (e.g. META-1 MeSH co-occurrences in MEDLINE) can be utilized as a powerful additional learning and training resources in machine-assisted indexing, classification and searching.

ISSN: 2324-9773

Doszkocs, T., & Sass, R. (1992). An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching. 3rd ASIS SIG/CR Classification Research Workshop, 12-36. doi:10.7152/acro.v3i1.12593

d. Search queries and incoming textual records, for instance new citations for MEDLINE or books to be cataloged and added to CATLINE, can be regarded as queries and can be subjected to the same automatic text analysis procedures as the database records themselves [NCBI (1991)].

e. Existing inverted file data structures, e.g. the many variants of B-trees used in a wide variety of operational retrieval systems , can support efficient weighted logic, closest match search strategies and ranked output of records, index terms and classification codes most similar to the query [Noreault (1977), Doszkocs (1980), Motzkin (1991)]. Both static and dynamic associative query and database processing can be supported in such a system.

f. Machine-assisted, rather than fully automatic procedures enable the system to maximally utilize the world knowledge and intelligence of human beings in recognizing and choosing appropriate items of interest that are "intelligently" filtered and suggested to them by the system. Such an approach is expected to synergistically combine the best abilities of the user with the prodigious associative memory and processing power of the system, and promote both consistency and cost effectiveness in indexing, classification and searching in a unified manner.

The most important associative data and information resources of the system are the **UMLS METATHESAURUS** knowledge base and the **ASSOCIATIVE PHRASE DICTIONARY**. These resources are indexed as typical inverted file databases.

The knowledge processing subsystem consists of three modules. The NLP (natural language subsystem) produces the associated phrases. The expert subsystem will handle rule-based search strategy modification. The neural network subsystem is under development. Its main purpose is to serve as a trained and supervised neural network of high-level MeSH classes based on the raw medical subject heading co- occurrence associations. Such a neural network can serve as a selection-filter from the pair-wise raw MeSH data associations in the context of particular queries.

Fig. 3 presents the main components of the system. Fig. 4 and Fig. 5 show examples of Metathesaurus subject heading associations with their frequency distributions and actual frequencies of co-occurrence.

ISSN: 2324-9773

Doszkocs, T., & Sass, R. (1992). An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching. 3rd ASIS SIG/CR Classification Research Workshop, 12-36. doi:10.7152/acro.v3i1.12593

Fig. 3 presents the main components of the system.

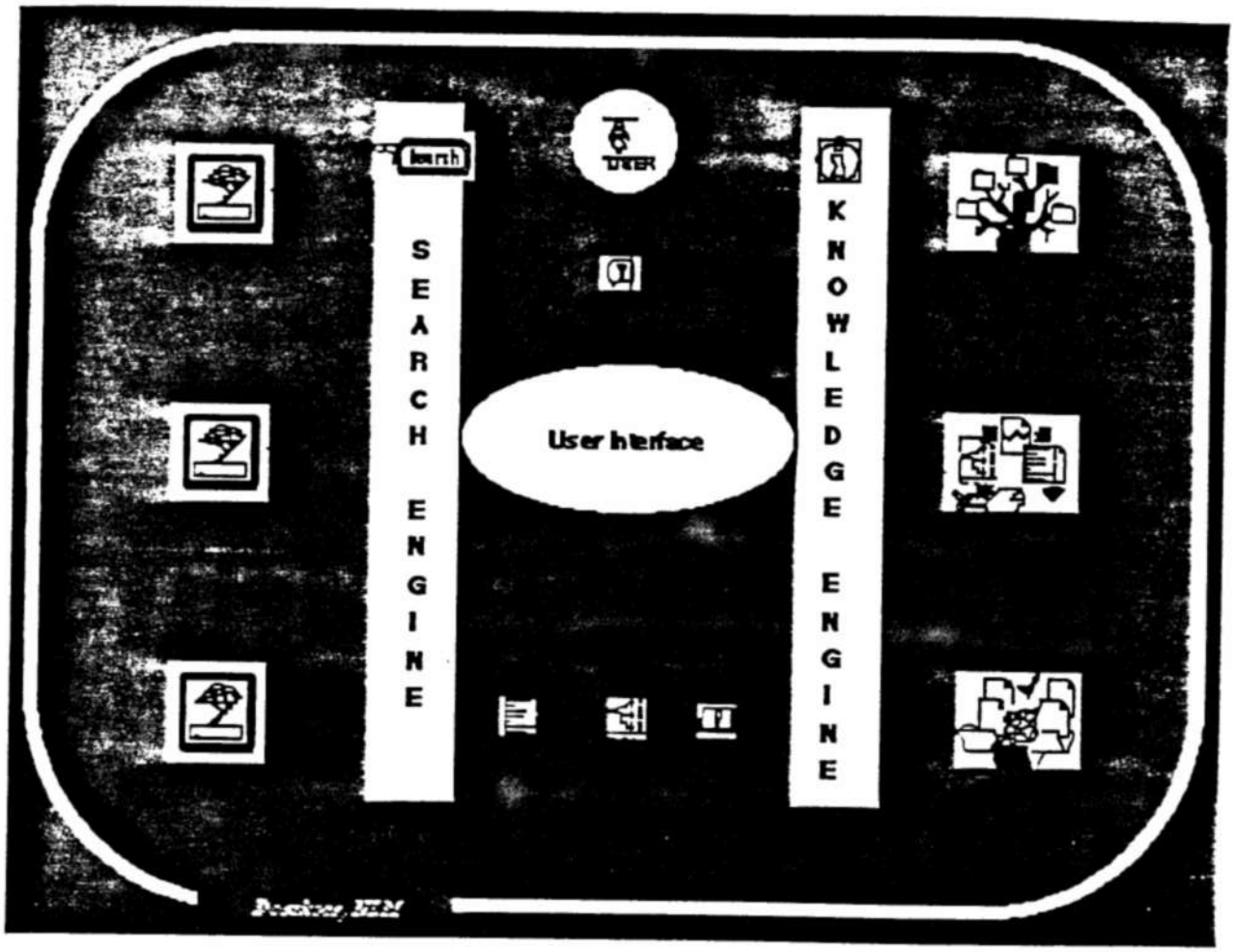
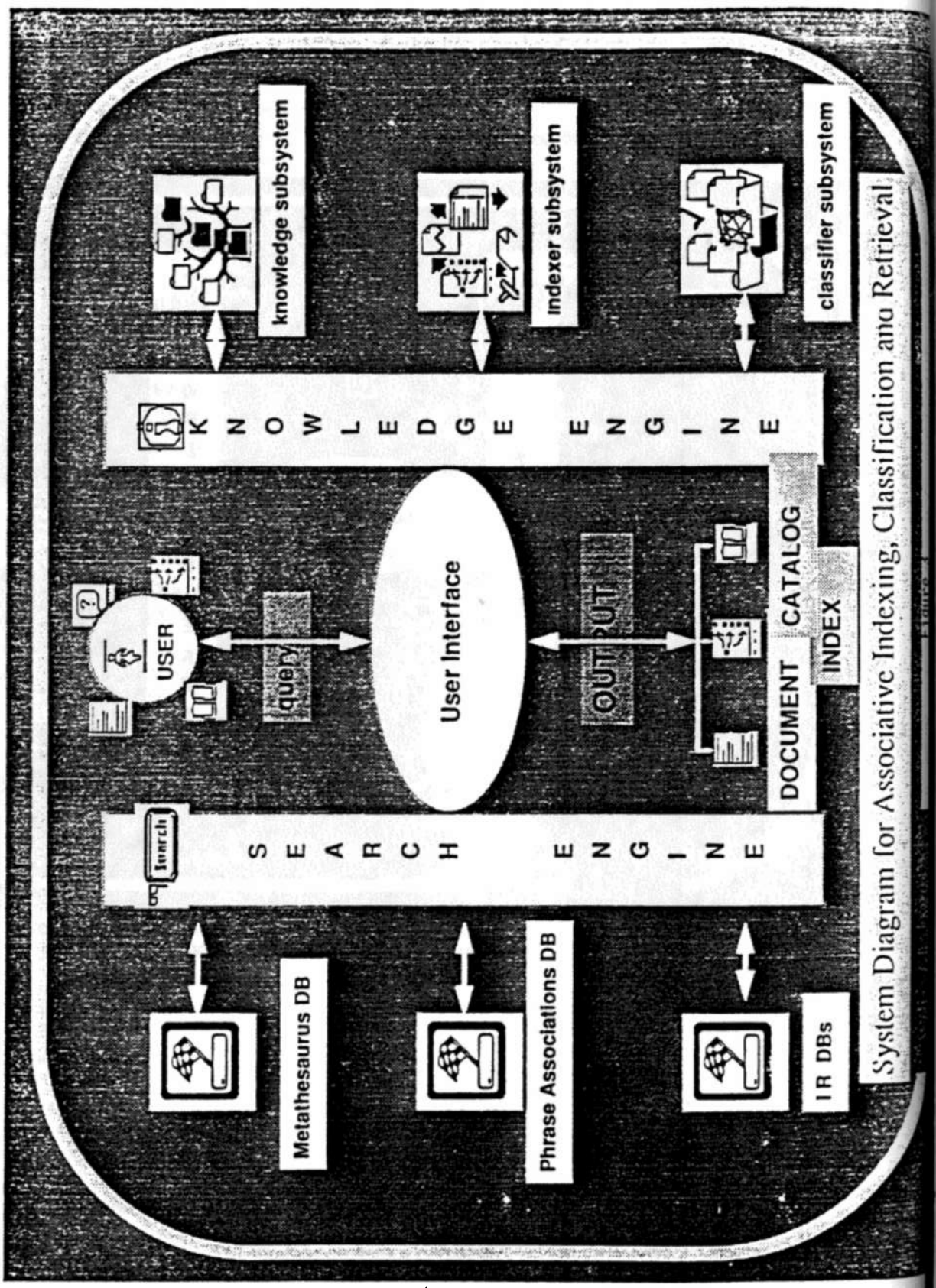


Figure 3. The Associative Indexing, Classification and Search System

ISSN: 2324-9773

Doszkocs, T., & Sass, R. (1992). An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching. 3rd ASIS SIG/CR Classification Research Workshop, 12-36. doi:10.7152/acro.v3i1.9996

Fig.3



ISSN: 2324-9773

Doszkocs, T., & Sass, R. (1992). An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching. 3rd ASIS SIG/CR Classification Research Workshop, 12-36. doi:10.7152/acro.v3i1.12593

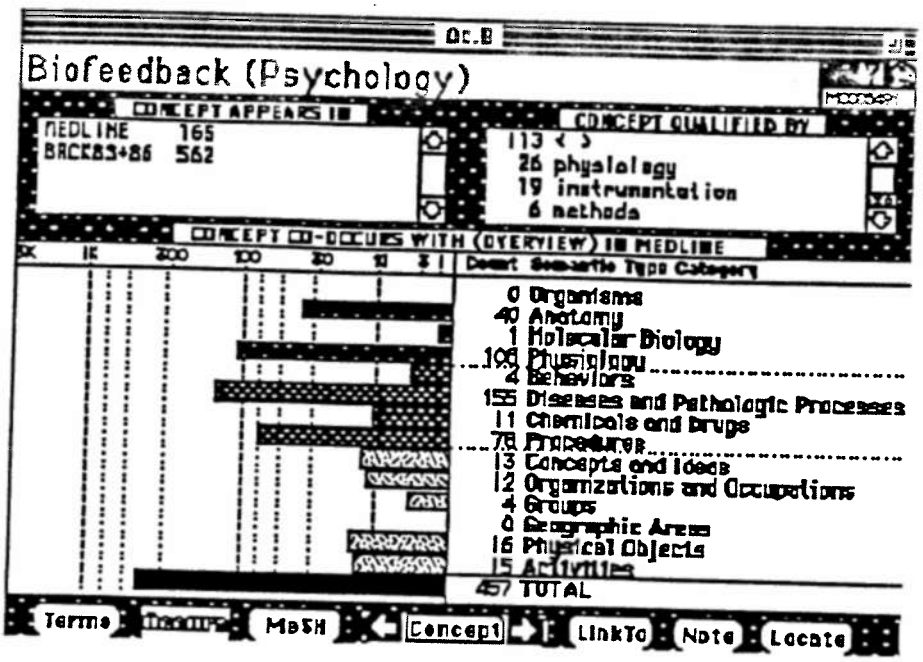


Figure 4. Frequency Distribution of Co-occuring MeSH in META-1

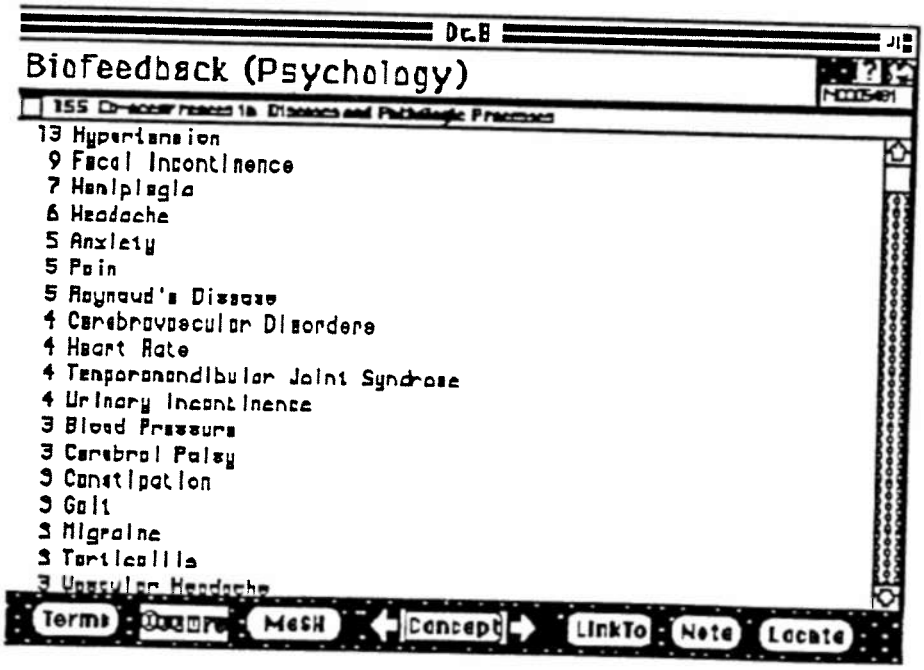


Figure 5. Medical Subject Heading Associations in META-1

ISSN: 2324-9773

Doszkocs, T., & Sass, R. (1992). An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching. 3rd ASIS SIG/CR Classification Research Workshop, 12-36. doi:10.7152/acro.v3i1.12593

Fig. 6 illustrates a typical free text phrase association record (with frequences of cooccurring phrases) derived from a set of MEDLINE database records retrieved by the medical subject heading BIOFEEDBACK (PSYCHOLOGY).

138	biofeedback
14	biofeedback training
2	EMG biofeedback training
8	thermal biofeedback
3	thermal biofeedback treatment
9	EMG biofeedback
2	EMG biofeedback studies
5	Electromyographic biofeedback
6	Biofeedback therapy
4	biofeedback-assisted relaxation
2	Temperature biofeedback
2	Biofeedback device
3	biofeedback-assisted control
2	functional biofeedback
2	quadriceps femoris muscle function following anterior cruciate ligament
28	patients
5	stroke patients
2	back pain patients
27	feedback
3	biological feedback
2	feedback stimuli
2	feedback signals
2	electromyographic feedback
2	feedback training
4	Behavioral treatment
16	incontinence
4	urinary incontinence
3	fecal incontinence
2	female incontinence
2	Faecal incontinence
15	hypertension
3	essential hypertension
2	arterial hypertension
13	headache
4	tension headache
4	vascular headache
2	migraine headache
2	Self-regulatory treatment
2	Autogenic training

Figure 6. Typical Phrase Association Record

Fig (7) displays a typical MEDLINE citation, with its title and abstract (phrases in bold have been automatically selected as potential noun phrases).

Wolf SL; Segal RL

Conditioning of the **spinal stretch reflex: implications** for **rehabilitation.**

Department of Rehabilitation Medicine, Emory University School of Medicine, Atlanta, GA 30322.
Phys Ther 1990 Oct;70(10):652-6

The **purpose** of this **article** is to describe a new **technique** that can potentially be **applied** to **patients** with **hyperactive spinal stretch reflexes** (SSRs). The **progression** of **clinical research** from **conditioning** of **individual muscles** or **muscle groups** (**electromyographic biofeedback**) to **conditioning SSRs** is explained. **Research data** from **subhuman primates** in addition to the first **human experiments** are reviewed. **Potential applications** of **SSR conditioning** are discussed, as are the issues requiring further **delineation** and **research** before the **specificity** of a **training effect** can be ascertained.

Figure 7. MEDLINE citation and abstract.

Fig. (8) shows a list of suggested medical subject headings retrieved by the closest match search engine from the **METATHESAURUS** medical subject heading co-occurrences and the **ASSOCIATIVE PHRASE DICTIONARY**. The results of the associative retrieval are to be evaluated and integrated by the indexing subsystem.

ISSN: 2324-9773

Phrases from the title and abstract	Topical MeSH identified by the Closest Match Search
PLEASE NOTE:	
(1) **bold face** Mesh headings were also assigned by an indexer	
(2) * MeSH headings were mapped from a title phrase and are presumed to be more significant	
(3) *italic MeSH headings could not be matched by any phrase*	
(4) shadowed MeSH headings were matched by associated phrases phrase	
	Cerebrovascular Disorders
clinical research	Clinical Research
Conditioning	*Conditioning
delineation	
electromyographic biofeedback	**Biofeedback (Psychology)** **Electromyography**
hyperactive spinal stretch reflexes	Hyperactivity Reflex ***Reflex, Stretch** ***Spine**
individual muscles	Muscles
muscle groups	Muscles
patients	Patients
Potential applications	
progression	
rehabilitation	***Rehabilitation** **Exercise Therapy**
Research data	**Research** Data Adjustment Data Reporting
specificity	Specificity
spinal stretch reflex (SSR)	**Reflex, Stretch**
SSR conditioning	Conditioning
subhuman primates	Primates
technique	
training effect	Training Technics

Figure 8. MeSH headings Matched by Closest-match Search Strategy and Phrase-to-MeSH Associations

ISSN: 2324-9773

Doszkocs, T., & Sass, R. (1992). An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching. 3rd ASIS SIG/CR Classification Research Workshop, 12-36. doi:10.7152/acro.v3i1.12593

Figure 9 shows the actual MeSH headings assigned to the sample citation by a human indexer, without the use of the machine-aided associative indexing subsystem. Please note that "exercise therapy" does not appear in the title and abstract, yet both the human indexer and the automatic associative display correctly included it among the chosen MeSH headings.

Biofeedback (Psychology)
Cerebrovascular Disorders
Electromyography
Exercise Therapy
Reflex, Stretch
Rehabilitation
Spine

Figure 9. MeSH Headings Actually Assigned by a Human Indexer

Examples of machine assisted classification and automatic search logic reformulation will be shown at the time of presenting this paper.

4. CONCLUSION

Research and development in information retrieval has provided the theoretical basis for a variety of associative information processing models and techniques over the past three decades.

While the operational feasibility and pragmatic utility of a number of statistical, linguistic and artificial intelligence approaches to fully automatic indexing, classification and searching has been demonstrated in recent years in actual, scaled-up, real life applications, many established, highly visible and heavily utilized systems of indispensable value, however, continue to rely on human intellectual effort in indexing, classification, and mediated searching of large interdisciplinary textual databases.

The proposed machine-assisted associative indexing, classification and searching approach is expected to capitalize on the considerable intellectual investment represented by the human indexing and classification effort traditionally applied to these databases. By combining "the best of both worlds", i.e. automatic indexing, classification and retrieval procedures with human judgement and selection, the experimental system promises to facilitate consistency, reduce cost,

ISSN: 2324-9773

add value and increase user satisfaction by its adaptation to the user's intelligence, desire for control, serendipity of recognition and the holism of the user-system synergy [Rice (1988)].

REFERENCES

Agosti, M.; Colotti, R.; Gradenigo, G. "A Two-Level Hypertext Retrieval Model for Legal Data", *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 316-325, 1991

Alberico, R.; Micco, M. *Expert Systems for Reference and Information Retrieval*. Westport, CT, Meckler, 1989.

Allen, B. "Topic Knowledge and Online Catalog Search Formulation", *Library Quarterly*, 61(2): 188-213, 1991

Bates, M.J. "Where Should the Person Stop and the Information Search Interface Start?", *Information Processing & Management*, 26(5): 575-591, 1990

Belkin, N.J.; Croft, W.B. "Retrieval Techniques", In: *Annual Review of Information Science and Technology*, 22: 109-145, 1987

Belkin, N.J.; Marchetti, P.G. "Determining the Functionality and features of an Intelligent Interface to an Information retrieval System", In" *ACM SIGIR 90*: 151-177, 1990

Biennier, F.; Pinon, J.M.; Guivarch, M. "A Connectionist Method to Retrieve Information in Hyperdocuments", In: *INNC 90 Paris, International Neural Network Conference*, 1: 444-448, Kluwer, Netherlands, 1990

Borgman, C.L. "All Users of Information retrieval Systems are not Created Equal: an Exploration into Individual Differences", *Information Processing and Management*, 26: 237-252, 1989

Brooks, H.M. "Expert Systems and Intelligent Information Retrieval". *Information Processing & Management*, 23(4): 367-382, 1987

Bush, V. "As We May Think", *Atlantic Monthly*, 176(1): 101-108, 1945

Chen, C. C. The Coming of Digital Visual Information Age: Implications for Information Access. In: *NIT '91: 4th International Conference on new Information Technology*, C.C. Chen (ed), pp. 5-18, Budapest, 1991

Cleverdon, C.W. "The Significance of the Cranfield tests of Index Languages", In: *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 3-12, 1991

Croft, W.B. "Approaches to Intelligent Information Retrieval." . *Information Processing & Management*, 23(4):249-254, 1987.

Croft, W.B.;; Turtle, H. R.; Lewis, D. D. "The Use of Phrases and Structured Queries in Information retrieval", In: *Proceedings of the Fourteenth Annual International ACM/ SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 32-45, 1991

Deerwester, S.; Dumais, S.T.; Furnas, G.W.; Landauer, T.K.; Harshman, R. "Indexing by Latent Semantic Indexing", *Journal of the American Society for Information Science*, 41(6): 391-407, 1990

Doszkocs, T.E. "AID, An associative Interactive Dictionary for Online Searching", *Online Review*, 2(2): 163-173, 1978

ISSN: 2324-9773

Doszkocs, T., & Sass, R. (1992). An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching. 3rd ASIS SIG/CR Classification Research Workshop, 12-36. doi:10.7152/acro.v3i1.12593

Doszkocs, T.E. AID, *An associative Interactive Dictionary for Online Bibliographic Searching*. (Doctoral Dissertation), University of Maryland, 124 pp, 1979. University Microfilms order: 79-25741]

Doszkocs, T.E.; Rapp, B.A. "Searching MEDLINE in English: a Prototype User Interface with Natural Language Query, Ranked Output, and relevance feedback", In: *Proceedings of the ASIS Annual Meeting*, 16: 131-139, 1979

Doszkocs, T.E.; Rapp, B.A.; Schoolman, H.M. "Automated Information Retrieval in Science and Technology", *Science*, 208: 25-30, 1980.

Doszkocs, T.E. "Implementing an Associative Search Interface in a Large Online Bibliographic Data Base Environment", In: *New Trends in Documentation and Information. Proceedings of the 39th FID Congress*, Sept. 25-28, 1978, Edinburgh, UK, pp. 295-297, Aslib, London, 1980.

Doszkocs, T.E. "CITE NLM: Natural-Language Searching in an Online Catalog", *Information Technology and Libraries*, 2(4): 364-380, 1983

Doszkocs, T.E. "Natural Language Processing in Information Retrieval", *Journal of the American Society for Information Science*, 37(4): 191-196, 1986

Doszkocs, T.E. , Reggia, J. ; Lin, X. Connectionist Models and Information Retrieval. In: *Annual Review of Information Science and Technology (ARIST)*, Volume 25, 209-260, 1990

Doszkocs, T,E. Neural Networks in Libraries: the Potential of a New Information Technology. In: *NIT '91: 4th International Conference on new Information Technology*, C.C. Chen (ed), pp. 27-33, Budapest, 1991

Doyle, L. B. Semantic Road Maps for Literature Searchers. *Journal of ACM*, 8, 553-578, 1961

Egan, D.E. "Individual Differences in Human-Computer Interaction", In: M. Helander (ed), *Handbook of Computer-Human Interaction*, The Netherlands, Elsevier, pp. 543-568, 1988

Egan, D.E.; Remde, J.R.; Gomez, L.M.; Landauer, T.K.; Eberhardt, J.; Lochbaum, C.C. "Formative Design Evaluation of SuperBook", *ACM Transactions on Information Systems*, 7: 30-57, 1989

Fowler, R.H.; Fowler, W.A.L.; Wilson, B. A. "Integrating Query, Thesaurus, and Documents through a Common Visual representation", *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 142-151, 1991

Fox, E.A. "Development of the CODER System: A testbed for Artificial Intelligence methods in Information Retrieval", *Information Processing and Management*, 23(4): 341-366, 1987

Fuhr, N.; Pfeifer, U. "combining Model-Oriented and Description-Oriented Approaches for Probabilistic Indexing", In: *Proceedings of the Fourteenth Annual International ACM/ SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 46-56, 1991

Gauch, S.; Smith, J.B. "An Expert System for Searching in Full-text". *Information Processing & Management*, 25(3): 253-263, 1989

Gibb, F.; Sharif, C. "CATALYST: An Expert Assistant for Cataloging", *Program*, 22: 62-71, 1988

Harman, D.; Benson, D.; Fitzpatrick, L.; Huntzinger, R.; Goldstein, C., "IRX: An Information Retrieval System for Experimentation and User Applications", *Proceedings of RIAO 88 Conference on User-Oriented Content-Based Text and Image Handling*, pp. 840-848, 1988

Hawkins, D.T. "Applications of Artificial Intelligence (AI) and Expert Systems for Online Searching", *Online*, 12(1): 31-43, 1988

ISSN: 2324-9773

Henry, H.K. "Human-Computer Interfaces and OPACS: Introductory Thoughts Related to INNOPAC", *Library Hi Tech*, 9(2): 63-68, 1991

Hersh, W.R.; Hickam, D. H. "A Comparative Analysis of Retrieval Effectiveness for Three methods of Indexing Aids-Related Abstracts", In: *Proceedings of the ASIS Annual Meeting*, 28: 211-217, 1991

Hjerppe, R.; Olander, B. "Cataloging and Expert Systems: AACR2 as a Knowledge Base", *Journal of the American Society for Information Science*, 40(1): 27-44, 1989

Humphrey, S.M. "MedInDex System: Medical Indexing Expert System", *Information Processing & Management*, 25(1): 73-88, 1989

Jones, R. L. "Automatic Document Content Analysis - the AIDA Project", to be published in *Library Hi Tech*, Fall 1991

Kwok, K.L. "Query Modification and Expansion in a Network with Adaptive Architecture", *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 192-199, 1991

Krulee, G.K.; Vrenios, A. "An Expert System Model of a Reference Librarian", *Library Software Review*, 8: 13-15, 1989

Larson, R. "Classification Clustering, Probabilistic Information retrieval, and the Online Catalog", *Library Quarterly*, 61(2): 133-173, 1991

Lelu, A. "Automatic Generation of "Hyper-Paths" in Information retrieval Systems: A Stochastic and an Incremental Algorithms", *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 326-335, 1991

Lesk, M.E. "Word-Word Associations in Document Retrieval Systems", *American Documentation*, 20: 27-38, 1969

Lesk, M. "The CORE Electronic Chemistry Library", *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 93-101, 1991

Levine, E. (moderator) "Full Text: from Tutorial to Innovations", In: *Proceedings of the ASIS Annual Meeting*, 28: 363-376 (misc. papers and abstracts), 1991

Lewis, D.D. "Learning in Intelligent Information Retrieval", *Machine Learning, Proceedings of the 8th International Workshop*, pp. 235-239, Morgan Kaufman, Santa Monica, CA, 1991

Lin, X. "A Self-Organizing Semantic map for Information Retrieval", *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 262-269, 1991

Lindberg, D.; Humphreys, B. "The UMLS Knowledge Sources: Tools for Building User Interfaces", *Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care*, 14: 121-125, 1990

Marchionini, G. "Information Seeking Strategies of Novices in a Full Text Electronic Encyclopedia", *Journal of the American Society for Information Science*, 29(3): 165-176, 1989

Marcus, R. S. "Computer and Human Understanding in Intelligent Retrieval Assistance", *Proceedings of the ASIS Annual Meeting*, 28: 349-59, 1991

Meadows, C.T.; Hewett, T.T.; Aversa, E.S. "A Computer Intermediary for Interactive Database Searching: 1. Design", *Journal of the American Society for Information Science*, 33(5): 325-332, 1982

ISSN: 2324-9773

Meadows, C.T.; Hewett, T.T.; Aversa, E.S. "A Computer Intermediary for Interactive Database Searching: 2. Evaluation", *Journal of the American Society for Information Science*, 33(6): 357-364, 1982

Micco, H.M.; Smith, I. "Expert Systems in Libraries: Do They Have a Place?", *Library Software Review*, 6: 125-28, 1987

Milstead, J.L. "Methodologies for Subject Analysis in Bibliographic Databases", *Information Processing & Management*, 28(3):407-431, 1992

Motzkin, D. "An Efficient Directory System for Document Retrieval", *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 291-304, 1991

Nash, Allen; Walkington, Toni, "Records Management Using Neural Networks", In: *AI in Industry & Government International Conference*, November 23-25, 1989, Hyderabad, India, pp. 575-585.

NCBI (National Center for Biotechnology Information), National Library of Medicine, Entrez: Sequences, PIR, GenBank, MEDLINE, *Compact Disc, October 1991, Pre-release 4 and Entrez User's Guide*.

Nelson, T. "No More teacher's Dirty looks", *Computer Decisions*, pp. 16-23, 1970

Nelson, T. Literary Machines, Swarthmore, PA, 1981, ISBN: 0-89347-052-X

Newby, G.B. "Navigation: A Fundamental Concept for Information Systems with Implications for Information retrieval", In: *Proceedings of the ASIS Annual Meeting*, 28: 111-117, 1991

Nielsen, J. Hypertext and Hypermedia, Academic Press, San Diego, 1990

Noreault, T.; Koll, M.; McGill, M.J. "Automatic Ranked Output from Boolean Searches in SIRE", *Journal of the ASIS*, 28(6): 333-339, 1977

Okaya, Y. "Terminology and Neural Network System: the Experience of the BRAIN and NLRI Project", In: *NIT '91: 4th International Conference on new Information technology*, C.C. Chen (ed), pp.133-143, Budapest, 1991

Partridge, D.; Wilks, Y. (eds), The Foundations of Artificial Intelligence, Cambridge University Press, Cambridge, 1990

Peat, H. J.; Willett, P. "The Limitations of Term Co-Occurrence Data for Query Expansion in Document Retrieval Systems", *Journal of the American Society for Information Science*, 42(5): 378-383, 1991

Pejtersen, A.M.; Olsen, S.E.; Zunde, P. "The Term Association Thesaurus: an Information Processing Aid Based on Associative Semantics", In: *Empirical Foundations of Information and Software Science III. Proceedings of the Third Symposium*, pp. 175-186, Plenum, New York, 1987

Pejtersen, A.M. "A Library System for Information Retrieval based on a Cognitive Task Analysis and Supported by an Icon-based interface", *Proceedings of the Twelfth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, edited by N.J. Belkin & C.J. Rijsbergen, SIGIR Forum 23(1): 40-47, 1989

Pollitt, A.S. "CANSEARCH: An Expert Systems Approach to Document retrieval", *Information Processing & Management*, 23(2): 119-138, 1987

Rau, L.F.; Jacobs, P.S. "Creating Segmented Databases from Free Text for Text Retrieval", *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 337-346, 1991

ISSN: 2324-9773

Reggia, J. A. "Self-Processing Networks and Their Biomedical Implications", *Proceedings of th... IEEE*, 76, 680-692, 1988

Rice, J. "Serendipity and Holism: the Beauty of OPACs", *Library Journal*: 138-140, February 15, 1988

Richardson, J. "Toward an Expert System in Reference Service: A Research Agenda for the 1990s", *College and Research Libraries*, 50(2): 231-248, 1989

Roysdon, C.; White, H.D., eds, Expert Systems in Reference Service, New York, Haworth Press, 1989

Salton, G. A Theory of Indexing, Philadelphia: Society for Industrial and Applied mathematics, 1975

Salton, G.; McGill, M.J. Introduction to Modern Information Retrieval, McGraw-Hill Book Company, New York, 1983

Salton, G. "On the Use of Term Associations in Automatic Information retrieval", In: *11th International Conference on Computational linguistics. Proceedings of COLING '86*, pp. 380-386, AKS, Bonn, Germany, 1986

Salton, G. Automatic Text Processing. The Transformation, Analysis, and Retrieval of Information by Computer. Reading, Massachusetts, Addison-Wesley, 530 pp., 1989

Sanchez, E.; Zadeh, L.A., eds. Approximate reasoning in Intelligent Systems, Decision, and Control. Oxford, Pergamon Press, 1987

Saracevic, T. "Individual Differences in Organizing, Searching and Retrieving Information", In: *Proceedings of the ASIS Annual Meeting*, 28: 82-86, 1991

Schneiderman, B. "Visual user Interfaces for Information Exploration", In: *Proceedings of the ASIS Annual Meeting*, 28: 379-384, 1991

Sharif, C.A.Y. Developing an Expert System for Classification of Books Using Micro-based Expert System Shells. *British Library Research Report 32*, London, British Library research and Development department, 1988

Smith, L.C. "Artificial Intelligence in Information Retrieval", *Annual Review of Information Science and Technology*, 22: 41-77, 1987

Sparck Jones, K. Automatic Keyword Classification for Information Retrieval, Butterworths, London, 1971

Sparck-Jones, K.; Kay, M. linguistics and Information Science, New York, Academic Press, 1973

Sparck Jones, K. Intelligent Interfaces for Information Retrieval Systems: Architecture Problems in the Construction of Expert Systems for Document Retrieval. In: *Future Trends in Information Science and Technology*, pp. 47-73, P.A. Yates, ed., Los Angeles, Taylor Graham, 1988

Stevens, M.E.; Guiliano, V.E.; Heilprin, L.B. Statistical Association methods for Mechanized Documentation, Washington, *National Bureau of Standards (Occasional Publication no. 269)*, 1965

Stiles, H. E. The Association Factor in Information Retrieval. *Journal of the Association for Computing Machinery*, 8(2): 271-279, 1961

Sutcliffe, R.F.E. "Distributed representations in a Text Based Information Retrieval System: a New Way of Using the Vector Space Model", *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 123-132, 1991

Turtle, H. Inference Networks for Document Retrieval, *Ph.d Thesis*, Computer and Information Science Department, University of Massachusetts at Amherst, 1990

ISSN: 2324-9773

Doszkocs, T., & Sass, R. (1992). An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching. 3rd ASIS SIG/CR Classification Research Workshop, 12-36. doi:10.7152/acro.v3i1.12593

van Rijsbergen, C.J. "A Theoretical Basis for the Use of Co-Occurrence Data in Information retrieval, *Journal of Documentation*, 33: 103-119, 1977

Vickery, A.; Brooks, H.; Robinson, B.; Vickery, B.C. "A Reference and Referral System Using Expert Systems Techniques", *Journal of Documentation*, 43(1): 1-23, 1987

Vizine-Goetz, D; Drabenstott, K.M. "Computer and manual Analysis of Subject terms Entered by Online Catalog Users", In: *Proceedings of the ASIS Annual Meeting*, 28: 156-161, 1991

Waltz, D. L. "IR Research and development at Thinking machines Corporation", In: *Proceedings of the ASIS Annual Meeting*, 28: 371-372, 1991

Warner, A.J.; Wenzel, P.H. "A Linguistic Analysis and categorization of Nominal Expressions", In: *Proceedings of the ASIS Annual Meeting*, 28: 186-191, 1991

Wilkinson, R.; Hingston, P. "Using the Cosine Measure in a Neural Network for Document Retrieval, *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, edited by A. Bookstein et al, Chicago, Illinois, 202-210, 1991

ISSN: 2324-9773

Doszkocs, T., & Sass, R. (1992). An Associative Semantic Network for Machine-Aided Indexing, Classification and Searching. 3rd ASIS SIG/CR Classification Research Workshop, 12-36. doi:10.7152/acro.v3i1.12593