

Division of Research
School of Business Administration

March 1988

**THE ADAPTIVE MAN-MACHINE NON-ARITHMETIC
INFORMATION PROCESSING SYSTEM REVISITED:
A PRIVATE DOCUMENT RETRIEVAL SYSTEM TO
FACILITATE QUERY REFINEMENT**

Working Paper #556
Manfred Kochen
Choon Y. Lee
Christopher Westland
The University of Michigan

FOR DISCUSSION PURPOSES ONLY

None of this material is to be quoted or
reproduced without the expressed permission
of the Division of Research

Copyright 1988
University of Michigan
School of Business Administration
Ann Arbor Michigan 48109

THE ADAPTIVE MAN-MACHINE NON-ARITHMETIC INFORMATION PROCESSING SYSTEM
REVISITED: A PRIVATE DOCUMENT RETRIEVAL SYSTEM TO FACILITATE QUERY REFINEMENT

Manfred Kochen, Choon Y. Lee, and Chris Westland
School of Business Administration
The University of Michigan
Ann Arbor, MI 48109

1. INTRODUCTION

AMNIPS, an acronym with the first letters of the system named in the title of this paper, was developed in 1960-3 at the IBM Thomas J. Watson Research Center in Yorktown as the main project of the information retrieval research program, headed by the first author. Other members of the project were E. Wong, H. Bohnert, C. Abraham, P. Reisner, D. Reich and F. Blair, with M. M. Flood serving as consultant. The original stimulus for the idea underlying AMNIPS came from a paper by Mary Stevens [1959]. AMNIPS was also influenced by a synthesis of ideas in artificial intelligence (Kochen [1960]), in which for the first time knowledge representation issues were articulated and emphasized. The system had been partially implemented by 1962 and was presented at a conference at that time (Kochen [1962], Kochen, et al., 1962)). Further elaborations were reported later (Kochen [1965]).

The central idea was to represent some of an individual's knowledge about a topic at a specified time by sentences of the form "name-predicate-name." "Non-arithmetic" in AMNIPS would today be called "symbolic." The predicates are all 2-place predicates selected by the system's document database (DDB) indexer, and so are the names. With the addition of variables, negation, conjunction and quantifiers, the set of possible sentences thus form a primitive, formal language, or an applied first-order predicate calculus with a time-varying terminal vocabulary. Its intended use was to extend its user's memory so as to help him with inferential reasoning and experience-modulated belief formation, recall, assimilation of new knowledge, and recognition of what he does not know, manifested as question-asking and query-clarification. It is an aspect of this latter function, focused more sharply on document retrieval, that we address here.

First, we will analyze AMNIPS in relation to some of the work in this area since 1962. Then we propose an extension of AMNIPS aimed at facilitating document based information retrieval (IR) by a user who does not know that content or objectives of the knowledge base about the document collection and who cannot at first accurately and clearly articulate his own information needs. Last, we show how such a system increases recall, precision and controls information overload.

2. USING AMNIPS AS A RELATIONAL KNOWLEDGE BASE MANAGER
WITH INFERENTIAL CAPABILITIES

Figure 1 presents a schematic of a typical IR system incorporating a knowledge base. The knowledge base represents DDB information content and

concept linkages in a highly compressed form. This research investigates the use of AMNIPS as an "inference engine" to automatically infer the user's actual information needs given the user's stated needs intersected with the limits of the database content as reflected in the knowledge base. At the same time, it performs as a relational knowledge base manager.

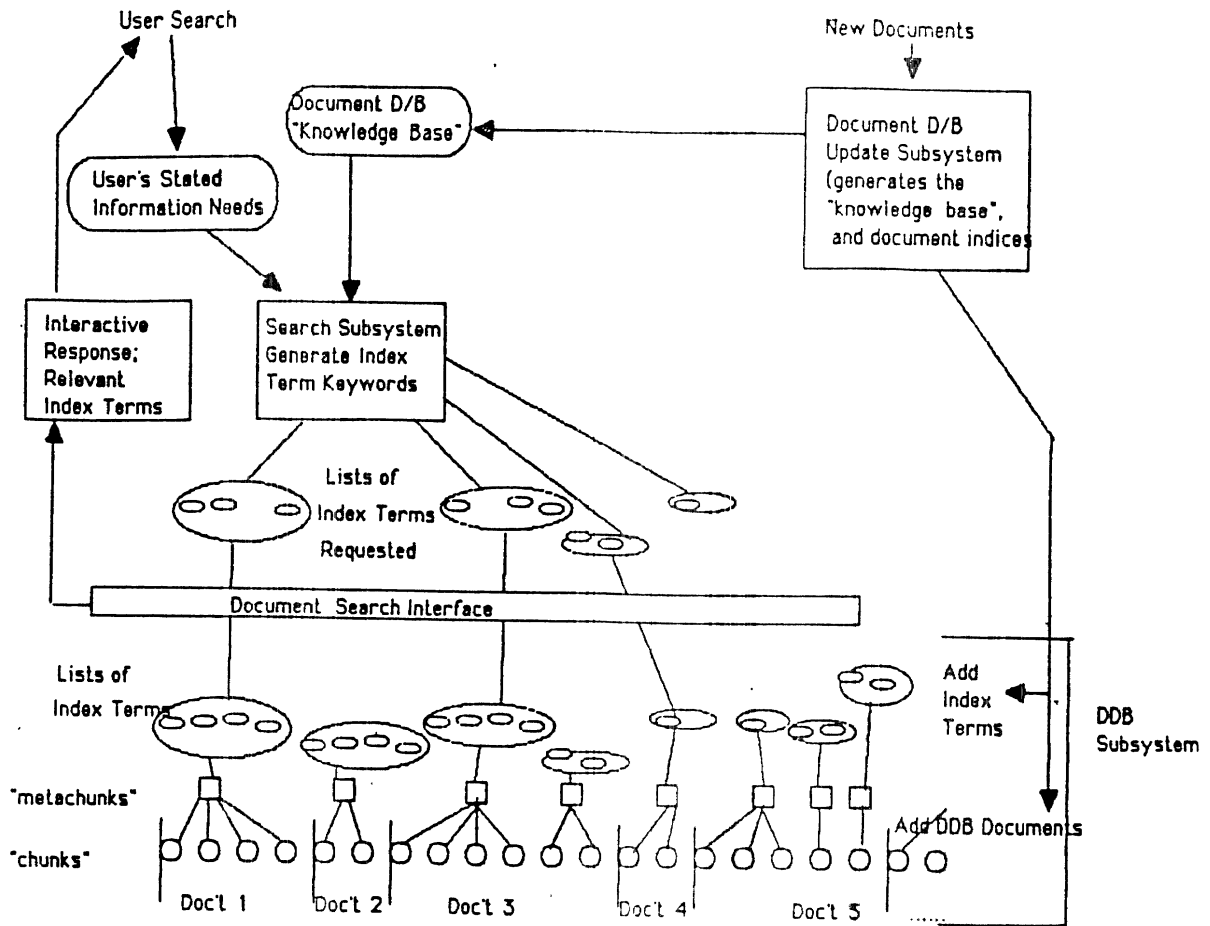


Figure 1: Schematic of a Generic Information Retrieval System

Numerous problems in business require the retrieval of information to satisfy poorly articulated requests. In a manufacturing business, production must continually respond to poorly understood and somewhat fickle market needs. In a professional business, e.g., medicine, accounting, or law, the professional is often required to make recommendations based on the professional's understanding of an extensive body of knowledge, in response to poorly articulated client needs. No professional can consistently keep current with even a fraction of the available knowledge in his field; thus automation of information retrieval is needed.

The AMNIPS algorithm has been adopted for this research because (1) it provides a simple inference engine, (2) it has been implemented in a "growing thesaurus" at IBM (Reisner [1965], Sandstrom, [1986]), and thus has been shown to be "practical," (3) it can represent the important concepts of "for some {argument}" and "for all {arguments}" (Woods [1975, p. 73]), and (4) it is representable to Church's lambda notation, which allows for extension to enrich the expressive capability of the language, and is suitable for LISP, PROLOG or other AI language processing. AMNIPS stores the knowledge of the DDB indexer, and combines it with the user's request to infer "concepts." The system's knowledge base may be expanded as the indexer assimilates new knowledge and the semantic network (SN) grows. Applied to document IR, the system

gets richer as the DDB accumulates more documents. In a sense, AMNIPS is a simulated mirror of the indexer's conceptual knowledge about DDB documents.

The system supports a user by extending his human limitations on storage, recall and knowledge of DDB content. The information stored in the system does not decay and is recalled in exactly the form in which it originally updated the DDB. Current research in statistical semantics may shed light on this area (Furnas, et al. [1983], Daniels [1986]) and shows the possibility for improving document retrieval by calculating simple statistical associations between an AMNIPS database and a general document database. This approach to document IR is somewhat different from what is generally called intelligent document retrieval (Heine [1986]), which tries to simulate the expertise of a human intermediary, i.e., the librarian.

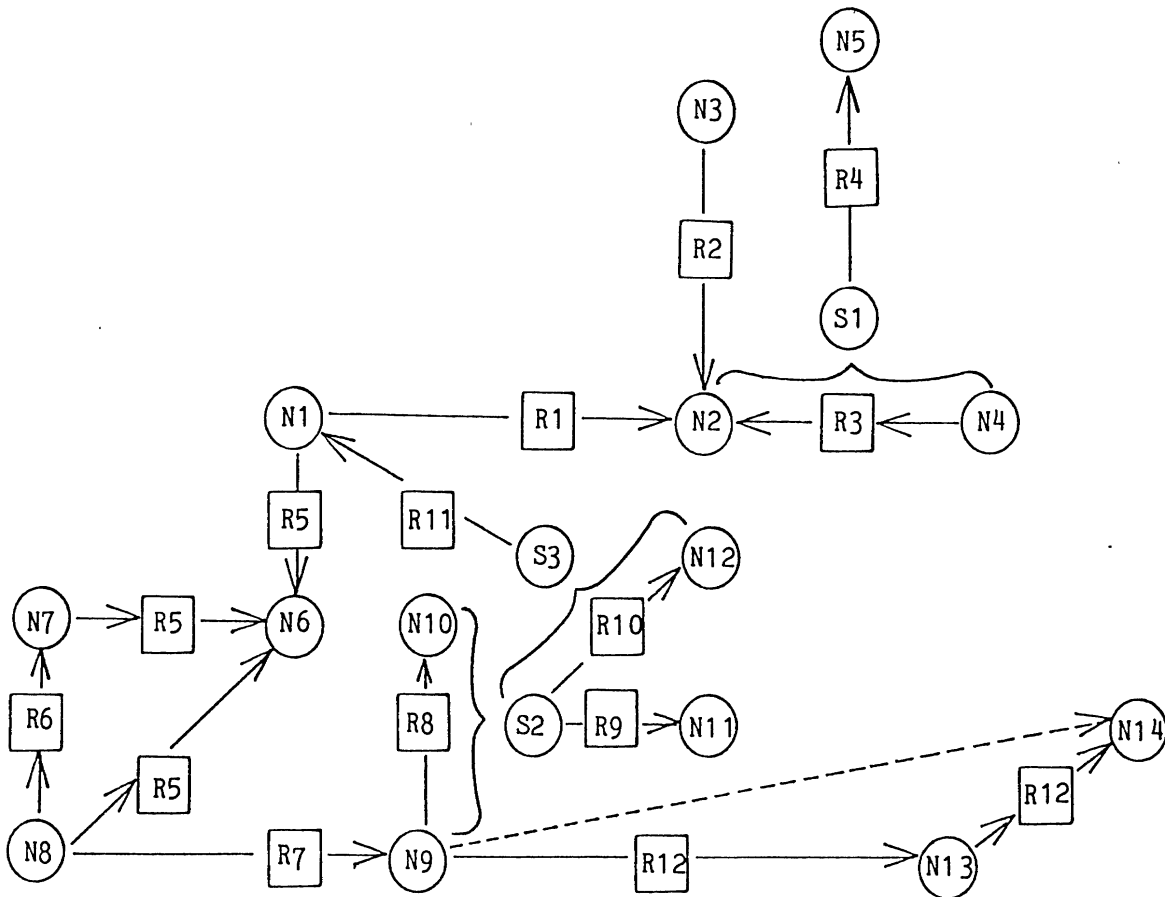
The two part predicate structure of AMNIPS defines a semantic net that provides information concerning (1) existence of correlation between keywords, (2) directionality, and (3) type of relationship. For example, the relationship "_____ is owned by _____" tells us that (1) we have an ownership relationship (vs., e.g., a causal relationship, or an attribute description) between the object and the subject, and (2) ownership has direction from the object to the subject. The subject and the direct object are individual names. The directionality of the predicate relation controls the construction of the SN. Unidirectional relations allow for the creation of hierarchies of "concepts." The SN processing returns lists of keyword values (which may be nouns, values, adjectives, and so forth) which reflect the logical intersection of user's information needs and information content on the DDB.

To fix ideas, consider a particular document based information retrieval system user with a personal document collection, DDB, who has at any time in his memory (primarily long-term) knowledge that he uses in recalling and retrieving items from the DDB when he needs them. Of course, the DDB and this memory are always changing. A snapshot of a tiny sample of the user's memory content, represented in AMNIPS language, is illustrated as a planar graph shown in Figure 2. One brief summary of this structure is that Chris Date (N1), the author of a widely purchased text (N2), is presenting a seminar that covers the relational model (N12) in the next six months (N11).

Since two-place relations are very limiting, to express such relations as when a predicate like sales volume applies, sentences, such as (N2, R3, N4) are named, e.g., S1 (upper right of Figure 2) and then inserted into the slots of other relations, e.g., R4. This can be iterated, as for S2 and S3. The predicates can, themselves, serve as names. Certain predicates, such as "is another name for", could be specified to be symmetric. Nearly every predicate has an inverse, such as "_____ is authored by _____" as the inverse of R1; a symmetric predicate is identical to its inverse. Other predicates, such as R12, can be specified to be transitive, so that (N9, R12, N14) can be immediately deduced completing the triangle with the dotted line at the lower right of Figure 2. If there are enough sentences like S1 and (N1, R1, N2), and so forth, in the knowledge base, AMNIPS can form inductive inferences like "all authors of popular books on databases are members of institutions located in California". In the context of Figure 2, "popular" must also have been related to sales of 225,000, "institutions" to "The Relational Institute" and "databases" to "The Relational Model".

AMNIPS-based document IR systems are not intended to support the organization of new documents. This must be the function of the indexer's judgment. This deficiency of AMNIPS is overcome by linking it with a general document retrieval system, which contains new documents and supports multiple users. Assume that a set of documents exists in both systems: AMNIPS represents the

indexer's interpretation of the documents but a general retrieval system represents the interpretation of a user group. The difference is the conceptual gap between the indexer and the user group. AMNIPS captures the



Legend:

- | | |
|---------------------------------|--|
| N1: Date, C. J. | R1: ___ is the author of ___ |
| N2: ISBN: | R2: ___ is the title of ___ |
| N3: An Introd. to DB Systems | R3: ___ is the nr. of copies sold of ___ |
| N4: 225000 | R4: ___ as of year ___ |
| N5: 1986 | R5: ___ is a member of ___ |
| N6: C&DCG | R6: ___ is another name for ___ |
| N7: Codd, E. F. | R7: ___ is the president of ___ |
| N8: Ted Codd | R8: ___ presents ___ |
| N9: The Relational Institute | R9: ___ during the period ___ |
| N10: TRI Seminars | R10: ___ about ___ |
| N11: January-June 1987 | R11: ___ by ___ |
| N12: The Relational Model | R12: ___ is located in ___ |
| N13: San Jose, California | |
| N14: Sample of a Knowledge Base | |

Sample of a Knowledge Base
Figure 2

conceptual difference, modifies the user's document retrieval query, and feeds it into the general system.

3. THE USE OF AMNIPS IN A DOCUMENT RETRIEVAL SYSTEM

Figure 1 depicts IR system-user interaction as an option on the document-search interface; i.e., the system may either use its suggested lists

of index terms to actually identify documents for retrieval, or it may return these lists to the user for his approval, and perhaps, modification. The interface is the IR system's automated librarian (AL), and offers the capability of user approval and modification of suggested requests, i.e., through user machine interaction. Interaction allows the user to refine and more clearly articulate his information needs, and it allows the IR search subsystem (via its knowledge base) to inform the user about information available on the DDB.

The AMNIPS "concept" list provides the document suggestions. The following sort of display for a database of Bread Recipe documents, on a recipes database, would be typical:

```

=====
|  User's Term   Related Terms   # of Documents |
|-----|-----|-----|
|  Flour        Flour           32             |
|              Corn Meal        9             |
|              Oat Meal         7             |
|  Sugar        Sugar           58             |
|              Honey            12            |
|              Aspartame        2             |
|              -----|-----|
|  Total Distinct Documents  92            |
|              -----|-----|
=====

```

Figure 3: Simulated CRT Display

In a typical IR session, the user submits a request (with the intention of retrieving quick bread receipts) $\{N_i\}; \{R_j\}$, e.g., $\{Flour, Sugar\}$ $\{ \text{_____} \text{ is ingredient with quick bread} \}$, and the automated librarian responds with $\{N'_i\}; \{R'_j\}$, e.g., $\{Flour, Corn Meal, Oat Meal, Sugar, Honey, Aspartame\}$; $\{ \text{_____} \text{ is ingredient in quick bread} \}$, representing "deep structure" concepts of $\{Mealy, Sweet\}$. The number of receipt documents containing the particular ingredient is indicated on the display. The user responds by selecting the suggested terms that are relevant to his real needs. Each iteration instructs the user on the content of the database. The effectiveness of this approach is suggested in Furnas, et al. [1983] and in the conclusion of Wang, Vanderdorpe, and Even [1985], along with the suggestion that at any iteration, the list of keywords be kept relatively short. As soon as the user completes the selection of term values, the automated librarian can again respond with suggestions, and the iterations can continue until the volume of documents falls below the "overload" limit.

4. HOW AMNIPS IMPROVES IR EFFECTIVENESS

The problem facing designers of IR system is how to devise an information retrieval system that allows an IR system's user to most effectively retrieve documents relevant to his request. Users are assumed not to know the content

of the document database, nor are they assumed to be able to accurately articulate their true information needs. Effective retrieval is determined by the standard performance measures of:

- Precision, which should be maximized
- -Recall, which should be maximized
- Requestors overload threshold, which should not be exceeded, and which is determined by user's bounded rationality and motivation.

Achievement of good Recall and Overload performance may be mutually incompatible.

AMNIPS stores the knowledge of an indexer of the DDB, and tries to form concepts from this knowledge set. The system's knowledge base is expanded as an indexer assimilates new knowledge and as the DDB grows. Applied to document retrieval, the system gets richer as an indexer gets more documents in his personal library. AMNIPS does not contain information about a document that has not been perused by an indexer. In this sense, AMNIPS is a simulated mirror of an individual's conceptual knowledge about his documents. The system supports a user by extending human limitations on storage and recall of information. The information stored in the system does not decay; it is recalled with perfect precision. This kind of approach for document retrieval is quite different from what is generally called intelligent document retrieval.

The automatic keyword list extension offered by AMNIPS provides the basic vehicle by which recall effectiveness is improved in the IR systems. As more related keywords are added to the list, the probability of retrieving relevant documents increases. Unfortunately, this occurs at the expense of precision and user overload. Control of precision and overload is accomplished in AMNIPS (1) by allowing interaction, to more closely identify the user's actual information needs to the IR system and (2) by carefully constructing a semantic net of index keywords and their relations. Each user request iteration allows AMNIPS to intersect more and more information (keywords and "concepts" from the query and the knowledge base) to elicit the user's actual information needs. The use of predicate relationships of even a few types (e.g., 10-20) has been shown by Fox [1981] to greatly enrich the explanatory power of SNs and to improve IR effectiveness. AMNIPS automates this explanatory power to control precision and recall.

BIBLIOGRAPHY

1. Daniels, P. J., "The User Modeling Function of an Intelligent Interface for Document Retrieval Systems," in Intelligent Information Systems for the Information Society, B. C. Brooks (ed.), Elsevier Science Publishers B. V. (North Holland), 1986, 162-176.
2. Fox, E., Lexical-Semantic Relations: Enhancing Effectiveness of Information Retrieval Systems. SIGIR Newsletter, March 1981.
3. Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T., Statistical Semantics: Analysis of the Potential Performance of Key-Word Information Systems. The Bell System Technical Journal, 62(6), July-August 1983, 1753-1806.
4. Heine, M. H., "Two Experiments on the Communication of Knowledge Through Database, in Intelligent Information Systems for the Information Society, B. C. Brooks (ed.), Amsterdam: North Holland, 1986, 121-140.
5. Kochen, M., Principles of Information Retrieval. Los Angeles: Melville, 1974.
6. Kochen, M., Cognitive Mechanisms, Report RAP-3. Yorktown Heights: IBM Research Center, 1960.
7. Kochen, M. Adaptive Mechanisms in Digital "Concept" Processing, Proc. Joint Autom. Control Cont., New York: IEEE, 1962, 50-59.
8. Kochen, M., C. Abraham, and E. Wong, "Adaptive, Man-Machine Concept Processing," IBM Final Report of Contract AF 19(604)-8446 for Electronic Research Directorate Air Force Cambridge Research Laboratories, June 1962.
9. Kochen, M., Some Problems in Information Science. Metuchen: Scarecrow, 1965.
10. Reisner, J., Semantic Diversity and a "Growing" Man-Machine Thesaurus, in Some Problems in Information Science, Kochen, M., Metuchen: Scarecrow, 1965.
11. Sandstrom, G., "Augmented Thesaurus for Multi-Contextual Descriptions," in Intelligent Information systems for the Information Society, B. C. Brooks (ed.), Amsterdam: North Holland, 1986, 192-210.
- 12c Stevens, M., A Machine Model for Recall. Proceeds from International Conference on Information Processing, Paris: UNESCO, 1959, 309-315.
13. Wang, Y., J. Vanderdorpe, and Evens, M., Relational Thesauri in Information Retrieval, JASIS, January 1985.
14. Woods, W. A., What's in a Link: Foundations for Semantic Networks, Representation and Understanding: Studies in Cognitive Science, Bobrow, D. G., and Collins, A. (eds.), New York: Academic Press, 1975.