REPORT R-493 OCTOBER, 1970 UILU-ENG 70-238

COORDINATED SCIENCE LABORATORY

# A DATA STRUCTURE FOR COGNITIVE INFORMATION RETRIEVAL

K.O. BISS
R. T. CHIEN
F. A. STAHL

UNIVERSITY OF ILLINOIS - URBANA, ILLINOIS

UILU-ENG 70-238

# A DATA STRUCTURE FOR COGNITIVE INFORMATION RETRIEVAL

by

K.O. Biss, R.T. Chien, F.A. Stahl

This work was supported in part by the Joint Services Electronics Program (U.S. Army, U.S. Navy & U.S. Air Force) under Contract DAAB 07-67-C-0199; and in part by Office of Education Grant OE-1-7-071213-4557.

# A DATA STRUCTURE FOR COGNITIVE INFORMATION RETRIEVAL

K.O. Biss, R.T. Chien, F.A. Stahl
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, Illinois

A new data structure developed in connection with a natural language question-answering system is described. The data structure is based upon a new high-order calculus. It allows a great degree of expressiveness and is suitable for extensive logical deduction.

The inadequacies of low-order schemes for the representation of natural language information are given. A formal description of the new high-order structure which overcomes these inadequacies is then presented along with the properties that make it more suitable and attractive for machine processing of natural language information.

# A DATA STRUCTURE FOR COGNITIVE INFORMATION RETRIEVAL

## I. Introduction

The storage and subsequent retrieval of large amounts of information has posed some very interesting problems about data structures and their utilization for efficient retrieval purposes. These problems become more acute when the system must deal with the answering of questions posed in natural language. This is due to the fact that a natural language question-answering system must be able to synthesize retrieved information in order to construct the answer to any given question. The internal structure of such systems, therefore, must reflect subtle semantic differences.

The purpose of this paper is to report a new data structure developed in connection with a natural language question-answering system. This data structure is based upon a new high-order calculus. It allows a great degree of expressiveness and is suitable for extensive logical deduction.

Early attempts to perform natural language question-answering were based in part upon the logical properties of the propositional calculus such as Darlington.[2] The DEACON [12] and PROSYNTHEX [11] systems relied more heavily upon the syntactic properties of natural language. There are two very well written surveys of natural language question-answering systems, both by Simmons.[8,9]

Due to the inadequacy of the propositional calculus as a data structure model for question-answering, a variety of other approaches using a limited number of predicates were reported.[1,4] Robinson [5] developed a semi-decision algorithm for the first-order predicate calculus and Green and Raphael [3] showed this to be a much more reasonable structure for representing and manipulating natural language information.

It became evident that even the first-order predicate calculus could not represent much of the information that is encountered in natural language. Robinson then developed a semi-decision procedure for a high-order representation.[6] The use of a high-order structure as the basic scheme for the representation of natural language information has far more potential than any of the lower-order schemes. One basic advantage in using the high-order representation lies in the fact that anything that can be represented or manipulated within a lower-order representation can be accomplished more efficiently in a high-order representation. In addition, the embedding properties of the high-order structure allow many things that were heretofore impossible to represent.

It is important to note that the concept of using a data structure that permits the embedding of information is not unique to the structure developed in this paper. Simmons, Burger, and Schwarcz [7,10] used a similar type representation scheme which was the best structure available for natural language question-answering until recently. However, the structure did not have the flexibility to allow powerful deduction.

The embedding properties of both of these approaches makes the use of lower-order structures less attractive for natural language question-answering systems. Motivation for the use of the new high-order structure based upon a logical calculus is given in Section II. Section III gives a formal definition of this high-order structure along with a number of annotated examples demonstrating its application.

## II. Structural Representation for Natural Language Information

Suppose that it is necessary to represent the sentence "John is in the crosswalk" in some formal structure. One such formal structure is the propositional calculus. We can represent this sentence in the propositional calculus by calling it the proposition

$$p \equiv \text{John is in the crosswalk}$$

The language of the propositional calculus consists of the propositional variables, p,q,r,... and the logical connectives $\sim$, $\wedge$, $\vee$, $\Rightarrow$. The sentences of this language are made up in the following way: the propositions and negations of propositions are sentences, and if P and Q are propositions then $P\wedge Q$, $P\vee Q$, $P\Rightarrow Q$ are sentences.

We can consider the propositions of this calculus in terms of a relational structure. Thus "John is in the crosswalk" can be written as the relation

$$p \equiv \text{in(John,crosswalk)}.$$

If we treat this sentence strictly as a proposition, then there is no way of comparing it to related sentences such as:

$$q \equiv \text{in(John,intersection)}$$

$$r \equiv \text{in(John,street)}$$

However, if these sentences are treated as true relations they can be compared so that the relationships that do exist can be recognized. These relationships may be used advantageously in data organization. For example, the data might be organized into property lists. For the example given above the property list of John might be:

John
in-crosswalk
⋮

The above structures are well defined and do allow a certain amount of deductive capability. Thus, if we are given the above property list, the rule

$$in(a,b) \wedge in(b,c) \Rightarrow in(a,c),$$

and the additional relation

$$in(crosswalk,intersection),$$

then it may be deduced that

$$in(John,intersection).$$

This structure is insufficient for natural language systems. This can be demonstrated with the following propositions

p = "Every boy is a person"

and

q = "John is a boy"

Within the propositional calculus it is impossible to deduce that John is a person from p and q. In order to be able to handle this type of deduction we must expand this structure to allow variables to appear in place of objects and allow quantification of these variables. In such a structure p might be represented by

$$\forall x(is(x,boy) \Rightarrow is(x,person)).$$

Then, knowing that

$$is(John,boy)$$

we can deduce

$$is(John,person).$$

A formal structure which does allow the use of variables ranging over objects and quantification over those variables is the first-order predicate calculus.

The language of the first-order predicate calculus is made up of predicate symbols P, Q, R,..., variables $x_1,x_2,\ldots$ which range over individuals, constant symbols, function symbols, the quantifiers $\forall$(for all) and $\exists$(there exists), and the logical operations $\sim$, $\wedge$, $\vee$, $\Rightarrow$. Terms are defined to be:

1. Constants and variables,
2. If $c_1\ldots c_n$ are terms and f is an n-ary function symbol then $fc_1\ldots c_n$ is a term.

The formulas of this language are made up in the following way:

1. If P is an n-ary predicate and $c_1,\ldots,c_n$ are terms then $Pc_1,\ldots,c_n$ is an atomic formula.
2. If A and B are formulas then $\sim A$, $A \wedge B$, $A \vee B$, and $A \Rightarrow B$ are formulas.
3. If A is a formula and x is a variable then $\forall xA$ and $\exists xA$ are formulas.
4. Nothing is a formula unless forced to be one by 1, 2, and 3.

Internal to a computer the sentences of the first-order predicate calculus, are represented as LISP expressions in Skolem prenex conjunctive normal form. Thus,

Every boy is a person

would be represented as

$$\forall x(\text{is}(x,\text{boy}) \Rightarrow \text{is}(x,\text{person}))$$

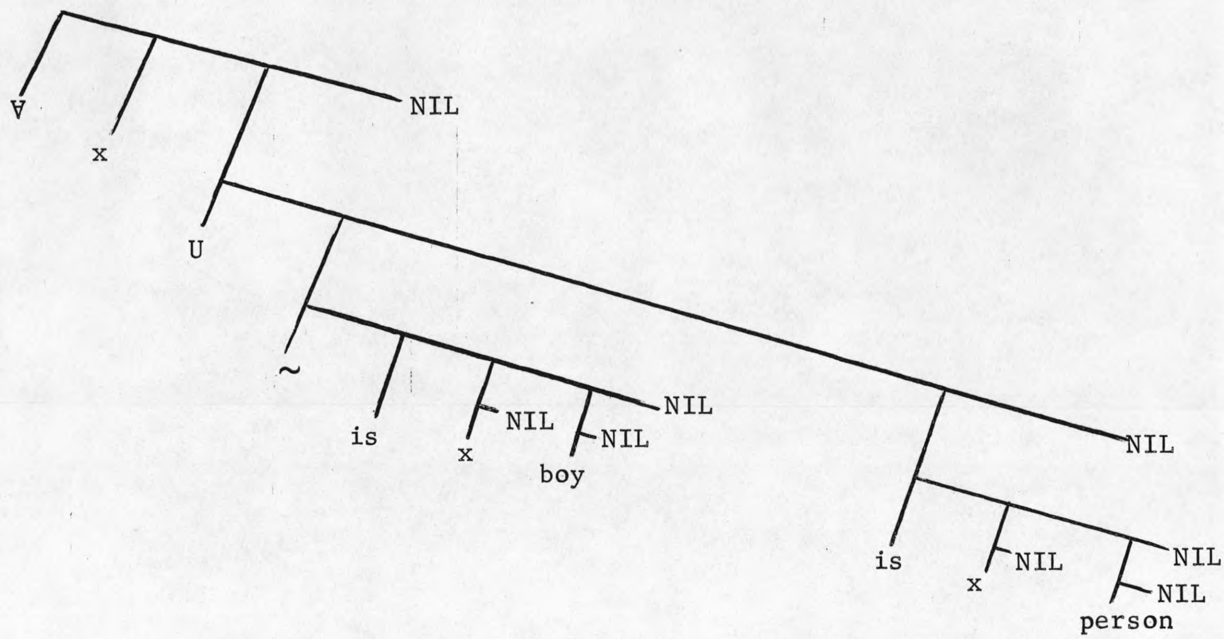
which in Skolem prenex conjunctive normal form would be

$$\forall x(\sim \text{is}(x,\text{boy}) \vee \text{is}(x,\text{person})).$$

The corresponding LISP expression is

```
(FA X (LOR (NEG IS (X) (BOY)) (IS (X) (PERSON))))
```

or equivalently




Even though the first-order predicate calculus is an improvement over the propositional and relational structures, it is still not powerful enough to be really useful in a natural language system. The main reasons for this are the inability to express relationships between relations, and to allow variables to range over relations as well as objects.

For example, suppose it is necessary to put into the first-order structure the sentences

John crossed the street after the light changed

or

A car must always yield to a pedestrian.

In the first case we are unable to put the sentence into the first-order structure because we have a relation, namely <u>after</u>, whose arguments are forced to be relations, namely crossed and changed, rather than some individuals. In the second case we cannot put the sentence into the first-order structure because we are faced with the quantification of a variable which ranges over situations not individuals. That is, the sentence states that for all possible situations a certain condition holds (i.e., that a car must yield to a pedestrian).

With the development given in this section one can see that the propositional, relational, and first-order structures have certain inadequacies in representing and manipulating natural language information. In the following section a high-order structure that overcomes these inadequacies is presented.

## III. A Data Structure Based on a High-Order Representation

For the reasons given in Section II we have gone to a higher-order structure for the representation of natural language information. In some ways the higher-order structure is similar to the first-order structure we defined previously. However, it is an extension of that structure in that we now allow relations to be embedded in other relations and we allow variables to range over these more complex structures. As a consequence, it is possible to represent situations as variables, the relationship between relations, and the modifications of terms. These features permit the representation of a wide range of natural language information. In addition, because of the generality of the structure chosen, the manipulation of information represented is greatly simplified.

A formal definition of this high-order structure will now be given: The following are defined within some natural language discourse $\mathcal{D}$.

1. $a_1$ is a <u>constant</u> iff $a_1$ is an object within $\mathcal{D}$.
2. $m_1$ is a <u>basic modifier</u> iff $m_1$ is a simple modifier within $\mathcal{D}$.
3. $c_1$ is a <u>modifying marker</u> iff $c_1$ indicates the occurrance of a modifier that is not simple in $\mathcal{D}$.

The high-order structure is made up of constant symbols $a_1, a_2, \ldots$, modifier symbols $m_1, m_2, \ldots$, modifying marker symbols $c_1, c_2, \ldots$, function symbols $f_1, f_2, \ldots$, variables that range over constants $x_1, x_2, \ldots$, n-ary relation symbols $P_1, P_2, \ldots$, variables that range over complex structures $y_1, y_2, \ldots$, and the logical symbols $\sim, \wedge, \vee, \Rightarrow$.

Terms are defined to be either:

1. constant symbols
2. all variables

3. complex structures, which are defined as either

   a. modified objects written $\bar{m}(\bar{a})$ where $\bar{m}$ is either a modifier, or a variable; and $\bar{a}$ is either a constant, a variable, or a complex structure (the interpretation of $\bar{m}(\bar{a})$ in $\mathfrak{D}$ is that $\bar{m}$ modifies $\bar{a}$). A modifier is either a basic modifier or, is of the form $c_1(\bar{b})$ where $c_1$ is a modifying marker and $\bar{b}$ is a constant or a modified object.

   b. n-ary relations over the terms $q_1,\ldots,q_n$ written $P(q_1,\ldots,q_n)$ where P is either an n-ary relation symbol or a variable which ranges over complex structures (this is interpreted to mean that $q,\ldots q_{n-1}$ and $q_n$ are in the relation P with each other).

4. if $t_1\ldots t_n$ are terms and f is an n-ary function symbol then $f(t_1\ldots t_n)$ is a term.

Formulas are defined in the following way:

1. n-ary relations and variables that range over n-ary relations are atomic formulas.
2. if A is an atomic formula and $\bar{m}$ is a modifier, then $\bar{m}(\bar{a})$ is a formula.
3. if A and B are formulas then $A \vee B$, $A \wedge B$, $\sim A$, $A \Rightarrow B$ are formulas.
4. if A is a formula and x is any variable then $\forall x(A)$ and $\exists x(A)$ are formulas.

For example, if the domain $\mathfrak{D}$ consists of English discourse relating to the operation of motor vehicles, then objects like car, driveway, and pedestrian are constants, and green, fast, and heavy are modifiers. Suppose

green car

appears in $\mathfrak{D}$. This would be represented as the modified object green(car)

because green is a modifier, car is a constant and green modifies car in 𝔇.

In order to appreciate this formalism think of green as a function whose value is equal to its argument with the additional property of being green. Thus, anything that applies to the object car, without qualification, also applies to the modified object green(car).

Similarly, if

the green car in the driveway

appears in 𝔇, then it would be represented by the modified object

(in(driveway))(green(car)).

This can be seen by first noting that in(driveway) is a modifier because in is a modifying marker and driveway is a constant. From the example above we know that green(car) is a modified object and a modifier followed by a modified object is itself a modified object. It also follows from the definition that

The green car in the driveway must yield to the pedestrian.

would be represented by the complex structure

must(yield((in(driveway))(green(car)),pedestrian)).

Notice, here, that yield is a binary relation between the objects

(in(driveway))(green(car))

and

pedestrian.

The binary relation yield is in turn modified by must.

Note that one might want to represent

(in(driveway))(car)

as the binary relation

in(car,driveway).

Using the new structure developed we can now represent the sentences which we could not handle previously. Thus, the sentence

A car must always yield to a pedestrian

would become

$$\forall_y (y \Rightarrow must(yield(car,pedestrian)))$$

where y is a variable that ranges over complex structures such as

(in(driveway))(car)

or

(in(crosswalk))(drunk(pedestrian)).

In the same manner, the sentence

John crossed the street after the light changed

would become

after(cross(John,street),change(light))

where _after_ is, in this example, an n-ary relation relating _cross_ and _change_. An algorithm for transforming natural language into this high-order structure is under development and will be discussed in a forthcoming paper.

The high-order structure described above permits the representation of all sentences which are representable in the lower-order structures (that is, the propositional, relational, and first-order structures). In addition to subsuming all lower-order structures, the high-order structure enables the representation of much information that could not be represented before. This increased generality makes it easier to represent natural language information. Furthermore, a deduction algorithm for the high-order structure has been developed. The algorithm is based upon Robinson's resolution principle. The details of this algorithm will be discussed in a forthcoming paper.

As an example of deductive capability within this high-order structure consider the sentence

A car must always yield to a pedestrian

and the question

Must a car in a driveway yield to a pedestrian in a crosswalk?

In the high-order representation scheme these would be

$$\forall y(y \Rightarrow must(yield(car,pedestrian))) \quad (1)$$

and

$$must(yield((in(driveway))(car),(in(crosswalk))(pedestrian))) \quad (2)$$

respectively. The answer to this question will be yes if it can be demonstrated that (2) is true. That is, by showing that it can be deduced from the information available in the system (namely (1) in this case). From (1) we can imply that

$$\forall y_1 \forall y_2(must(yield(y_1(car),y_2(pedestrian))) \quad (3)$$

must hold.

To see this, note that $y_1$ and $y_2$ are modifiers of car and pedestrian, respectively. That means that a car in all situations $y_1$ must yield to a pedestrian in all situations $y_2$, but that's just what (1) says when restricted to situations dealing with cars and pedestrians.

Now, consider the following assignment for variables $y_1$ and $y_2$ in (3):

$y_1 \equiv$ (in(driveway)), and
$y_2 \equiv$ (in(crosswalk))

Then (3) would appear as:

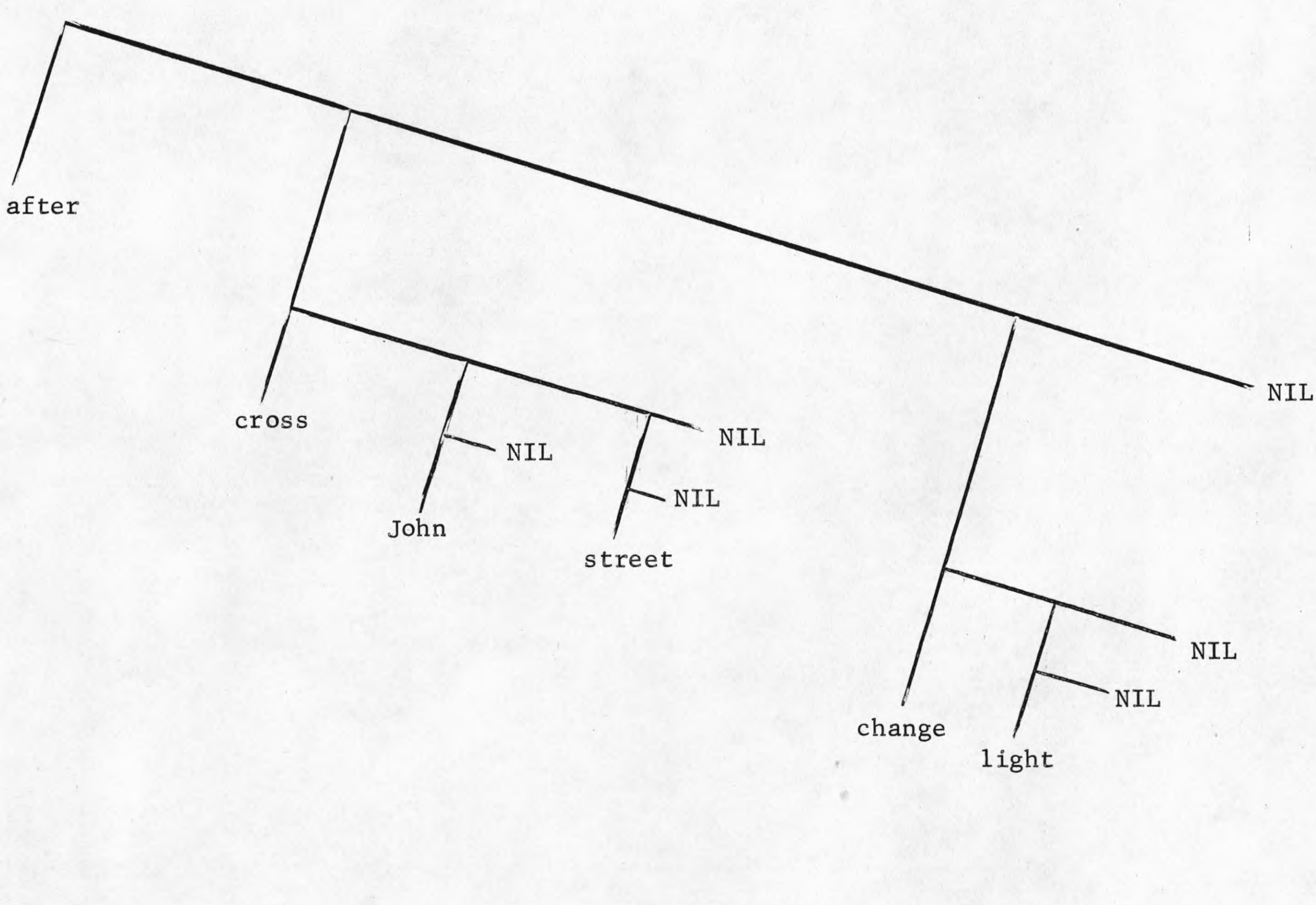
$$must(yield((in(driveway))(car),(in(crosswalk))(pedestrian))) \quad (4)$$

But (4) is exactly what we set out to deduce, so the answer would be yes.

Internally the new structure can be considered in certain respects as an extension of the first-order structure. Thus, where only objects appeared in the LISP expressions in the first-order structure, complex structures may now appear. The sentence "John crossed the street after the light changed" would appear as the LISP expression:

(AFTER (CROSS (JOHN) (STREET)) (CHANGE (LIGHT)))

or equivalently



## IV. Conclusion

In any high quality natural language question-answering system one must have:

1. an internal data structure sufficiently rich to represent natural language information.
2. a method of transforming natural language into that structure.
3. a strong deduction algorithm for manipulating the information in that structure.

It should be pointed out that these are not separate problems. In fact, the data structure plays the central role in such a system for the two following reasons. First, the realization of a transformational algorithm is completely dependent upon the characteristics of the internal structure. Secondly, the deduction algorithm can only be as powerful as the expressiveness of the internal data structure.

In this paper we have discussed the inadequacies of low-order schemes for the representation of information in natural language systems. A new high-order structure for the representation of natural language information was then presented. It has been shown that within this new structure a much wider range of natural language information can be represented. The powerfulness of representation of the high-order structure is due to its embedding property and its ability to allow quantification over complex structures.

Extensive investigation indicates that this high-order structure is extremely well suited for both the transformation of natural language information and powerful deduction, thus making it desirable for high quality natural language question-answering systems.

REFERENCES

[1] Black, F. S., "A Deductive Question-Answering System," in Minsky, M., (ed.) Semantic Information Processing, MIT Press, Cambridge, Mass., 1968, pp. 354-402.

[2] Darlington, J. L., "Translating Ordinary Language into Symbolic Logic," MAC M-149, Project MAC Memo, Mass. Inst. of Tech., Cambridge, Mass., March, 1964.

[3] Green, C. C. and Raphael, B., "The Use of Theorem Proving Techniques in Question Answering Systems," Proceedings of the ACM National Conference, 1968, pp. 169-181.

[4] Raphael, B., "A Computer Program for Semantic Information Retrieval," in Minsky, M., (ed.) Semantic Information Processing, MIT Press, Cambridge, Mass., 1968, pp. 33-145.

[5] Robinson, J. A., "A Machine-Oriented Logic Based on the Resolution Principle," Journal of the ACM, January, 1965, pp. 23-41.

[6] Robinson, J. A., "Mechanizing Higher-Order Logic," in Meltzer, B., and Michie, D., Machine Intelligence 4, American Elsevier, 1969, pp. 151-170.

[7] Schwarcz, R. M., Burger, J. F., and Simmons, R. F., "A Deductive Question-Answerer for Natural Language Inference," Communications of the ACM, March, 1970, pp. 167-183.

[8] Simmons, R. F., "Answering English Questions by Computer: A Survey," Communications of the ACM, Vol. 8, No. 1, January, 1965, pp. 53-69.

[9] Simmons, R. F., "Natural Language Question-Answering Systems: 1969," Communications of the ACM, Vol. 13, No. 1, January, 1970, pp. 15-36.

[10] Simmons, R. F., Burger, J. F., and Schwarcz, R. M., "A Computational Model of Verbal Understanding," AFIPS, FJCC, Spartan Books, New York, 1968, pp. 441-456.

[11] Simmons, R. F., Klein, S., and McConogue, K. L., "Indexing and Dependency Logic for Answering English Questions," American Documentation, Vol. 15, No. 3, 1964, pp. 196-204.

[12] Thompson, F. B. et al., "DEACON Breadboard Summary," RM 64TMP-9, TEMPO General Electric Company, Santa Barbara, Calif., 1964.

## Distribution List as of August 1, 1970

ESD (ESTI)
L. G. Hanscom Field
Bedford, Mass 01731 2 Copies

Mr I. A. Balton
Institute for Exploratory Research
Code: AMSEL-XL
U. S. Army Electronics Command
Fort Monmouth, New Jersey 07703

LTC Howard W. Jackson
Deputy Dir of Electr. & Solid St. Sciences
Air Force Office of Scientific Research
1400 Wilson Boulevard
Arlington, Virginia 22209 5 Copies

Defense Documentation Center
Attn: DDC-TCA
Cameron Station
Alexandria, Virginia 22314 50 Copies

Director, Electronics Program
Attn: Code 427
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217 3 Copies

Naval Air Systems Command
AIR 03
Washington, D. C. 20360 2 Copies

Naval Electronic Systems Command
ELEX 03, Room 2046 Munitions Building
Department of the Navy
Washington, D.C. 20360 2 Copies

Director
Naval Research Laboratory
Washington, D.C. 20390
Attn: Code 2027 6 Copies

Commander
U. S. Naval Ordnance Laboratory
Attn: Librarian
White Oak, Md. 20910 2 Copies

Commanding General
Attn: STEWS-RE-L, Technical Library
White Sands Missile Range
New Mexico 88002 2 Copies

Commander
Naval Electronics Laboratory Center
Attn: Library
San Diego, Calif 92152 2 Copies

Dr L. M. Hollingsworth
AFCRL (CRN)
L. G. Hanscom Field
Bedford, Massachusetts 01731

Division of Engineering & Applied Physics
210 Pierce Hall
Harvard University
Cambridge, Massachusetts, 02138

Director
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Miss R. Joyce Harman
Project MAC, Room 810
545 Technology Square
Cambridge, Mass 02139

Professor R. H. Rediker
Electrical Engineering Prof.
Mass. Institute of Technology
Building 13-3050
Cambridge, Mass 02139

Raytheon Company
Research Division Library
28 Seyon Street
Waltham, Massachusetts 02154

Sylvania Electronic Systems
Applied Research Laboratory
Attn: Documents Librarian
40 Sylvan Road
Waltham, Mass 02154

Commanding Officer
Army Materials & Mechanics Res. Center
Attn: Dr H. Priest
Watertown Arsenal
Watertown, Massachusetts 02172

MIT Lincoln Laboratory
Attn: Library A-082
P.O. Box 73
Lexington, Mass. 02173

Commanding Officer
Office of Naval Research Branch Office
495 Summer Street
Boston, Massachusetts 02210

Commanding Officer (Code 2064)
U. S. Naval Underwater Sound Laboratory
Fort Trumbull
New London, Connecticut 06320

Dept of Eng & Applied Science
Yale University
New Haven, Conn 06520

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-CT-A
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-CT-D
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-CT-I
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-CT-L (Dr W. S. McAfee)
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-CT-O
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-CT-R
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Fort Monmouth, New Jersey 07703
Attn: AMSEL-CT-S

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-DL
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-GG-DD
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-KL-D
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-KL-E
Fort Monmouth, New Jersey 07707

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-KL-I
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-KL-SM
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-KL-S
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-KL-T
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-NL-A
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-NL-C
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-NL-D (Dr H. Bennett)
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-NL-P
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-SC
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-VL-D
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-VL-F
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-WL-D
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-XL-DT
Fort Monmouth, New Jersey 07703

Commanding General
U. S. Army Electronics Command
Attn: AMSEL-XL-D
Fort Monmouth, New Jersey 07703

Mr Norman J. Field, AMSEL-RD-S
Chief, Office of Science & Technology
Research and Development Directorate
U. S. Army Electronics Command
Fort Monmouth, New Jersey 07703

Project Manager
Common Positioning & Navigation Systems
Attn: Harold H. Bahr (AMCPM-NS-TM),
Bldg. 439
U. S. Army Electronics Command
Fort Monmouth, New Jersey 07703

U. S. Army Munitions Command
Attn: Science & Technology
Info Br., Bldg 59
Picatinny Arsenal, SMUPA-RT-S
Dover, New Jersey 07801

European Office of Aerospace Research
Technical Information Office
Box 14, FPO New York 09510

Director
Columbia Radiation Laboratory
Columbia University
538 West 120th St.
New York, N.Y. 10027

New York University
Engineering Library
Bronx, New York 10453

Mr Jerome Fox, Research Coordinator
Polytechnic Institute of Brooklyn
333 Jay St.
Brooklyn, N. Y. 11201

Airborne Instruments Laboratory
Deerpark, New York 11729

Dr. W. R. Lepage, Chairman
Syracuse University
Dept of Electrical Engineering
Syracuse, N. Y. 13210

Rome Air Development Center
Attn: Documents Library (EMTLD)
Griffiss Air Force Base, N. Y. 13440

Mr H. E. Webb (EMBIS)
Rome Air Development Center
Griffiss Air Force Base, N.Y. 13440

Professor James A. Cadzow
Department of Electrical Engineering
State University of New York at Buffalo
Buffalo, N. Y. 14214

Dr. A. G. Jordan
Head of Dept of Elec Engineering
Carnegie-Mellon University
Pittsburgh, Penn 15213

Hunt Library
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA. 15213

Lehigh University
Dept of Electrical Engineering
Bethelehem, Pennsylvania 18015

Commander (ADL)
Naval Air Development Center
Attn: NADC Library
Johnsville, Warminister, Pa 18974

Technical Director (SMUFA-A2000-107-1)
Frankford Arsenal
Philadelphia, Pennsylvania 19137

Mr M. Zane Thornton, Chief, Network
Engineering, Communications and
Operations Branch, Lister Hill
National Center/ Biomedical Communications
8600 Rockville Pike
Bethesda, Maryland 20014

U. S. Post Office Dept
Library- Room 6012
12th & Pennsylvania Ave. N.W.
Washington, D.C. 20260

Technical Library
DDR&E
Room 3C-122, The Pentagon
Washington, D.C. 20301

Director for Materials Sciences
Advanced Research Projects Agency
Department of Defense
Washington, D.C. 20301

Assistant Director, (Research)
Office of Director of Defense Research
& Engineering
Pentagon, Rm 3C128
Washington, D.C. 20301

Chief, R & D Division (340)
Defense Communications Agency
Washington, D.C. 20305

Commanding General
U. S. Army Materiel Command
Attn: AMCRD-TP
Washington, D.C. 20315

Director, U. S. Army Material
Concepts Agency
Washington, D.C. 20315

Hq USAF (AFRDD)
The Pentagon
Washington, D.C. 20330

Hq USAF (AFRDDG)
The Pentagon
Washington, D.C. 02330

Hq USAF (AFRDSD)
The Pentagon
Washington, D.C. 20330

AFSC (SCTSE)
Andrews Air Force Base, Maryland 20331

Dr I. R. Mirman
Hq AFSC (SGGP)
Andrews AFB, Maryland 20331

Naval Ship Systems Command
Ship 031
Washington, D.C. 20360

Naval Ship Systems Command
Ship 035
Washington, D.C. 20360

Commander
U. S. Naval Security Group Command
Attn: G43
3801 Nebraska Avenue
Washington, D.C. 20390

U. S. Naval Oceanographic Office
Attn: M. Rogofsky, Librarian (Code 640)
Washington, D.C. 20390

Director
Naval Research Laboratory
Washington D.C. 20390
Attn: Dr A. Brodizinsky, Sup. Elec Div

Director
Naval Research Laboratory
Washington, D.C. 20390
Attn: Maury Center Library (Code 8050)

Director
Naval Research Laboratory
Washington, D.C. 20390
Attn: Dr W. C. Hall, Code 7000

Director
Naval Research Laboratory
Attn: Library, Code 2029 (ONRL)
Washington, D.C. 20390

Dr G.M.R. Winkler
Director, Time Service Division
U. S. Naval Observatory
Washington, D.C. 20390

Colonel E. P. Gaines, Jr
ACDA/FO
1901 Pennsylvania Ave. N. W.
Washington, D.C. 20451

Commanding Officer
Harry Diamond Laboratories
Attn: Mr Berthold Altman (AMXD)-TI)
Connecticut Ave & Van Ness St., N.W.
Washington, D.C. 20438

Central Intelligence Agency
Attn: CRS/ADD Publications
Washington, D.C. 20505

Dr H. Harrison, Code RRE
Chief, Electrophysics Branch
National Aeronautics & Space Admin.
Washington, D.C. 20546

The John Hopkins University
Applied Physics Laboratory
Attn: Document Librarian
8621 Georgia Avenue
Silver Spring, Maryland 20910

Commanding Officer (AMXRD-BAT)
U.S. Army Ballistics Research Laboratory
Aberdeen Proving Ground
Aberdeen, Maryland 21005

Technical Director
U. S. Army Land Warfare Laboratory
Aberdeen Proving Ground
Aberdeen, Maryland 21005

Electromagnetic Compatibility
Analysis Center (ECAC)
Attn: (ACOAT)
North Severn
Annapolis, Maryland 21402

Commanding Officer
U. S. Army Engineer Topographic Labs
Attn: STINFLO Center
Fort Belvoir, Virginia 22060

Director (NV-D)
Night Vision Laboratory, USAECOM
Fort Belvoir, Virginia 22060

U. S. Army Mobility Equipment Research
and Development Center
Attn: Technical Document Center
Bldg 315
Fort Belvoir, Virginia 22060

Dr Alvin D. Schnitzler
Institute for Defense Analyses
Science and Technology Division
400 Army-Navy Drive
Arlington, Virginia 22202

Director, Physical & Eng. Sciences Div.
3045 Columbia Pike
Arlington, Virginia 22204

Commanding General
U. S. Army Security Agency
Attn: IARD-T
Arlington Hall Station
Arlington, Virginia 22212

Dr Joel Trimble, Code 437
Information Systems Branch
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Commanding General
USACDC Institute of Land Combat
Attn: Technical Library, rm 636
2461 Eisenhower Avenue
Alexandria, Virginia 22314

VELA Seismological Center
300 North Washington St.
Alexandria, Virginia 22314
U. S. Naval Weapons Laboratory
Dahlgren, Virginia 22448

Research Laboratories for the Eng.
Sciences, School of Engineering &
Applied Science
University of Virginia
Charlottesville, Va. 22903

Dr Herman Robl
Deputy Chief Scientist
U. S. Army Research Office (Durham)
Box CM, Duke Station
Durham, North Carolina 27706

Richard O. Ulsh (CRDARD-IP)
U. S. Army Research Office (Durham)
Box CM, Duke Station
Durham, North Carolina 27706

ADTC (ADBPS-12)
Eglin AFB, Florida 32542

Commanding Officer
Naval Training Device Center
Orlando, Florida 32813

Technical Library, AFETR
(ETV,MU-135)
Patrick AFB, Florida 32925

Commanding General
U. S. Army Missile Command
Attn: AMSMI-RR
Redstone Arsenal, Alabama 35809

Redstone Scientific Information Center
Attn: Chief, Document Section
U. S. Army Missile Command
Redstone Arsenal, Alabama 35809

AUL3T-9663
Maxwell AFB, Alabama 36112

Hq AEDC (AETS)
Attn: Library/Documents
Arnold AFS, Tennessee 37389

Case Institute of Technology
Engineering Division
University Circle
Cleveland, Ohio 44106

NASA Lewis Research Center
Attn: Library
21000 Brookpark Road
Cleveland, Ohio 44135

Director
Air Force Avionics Laboratory
Wright-Patterson AFB, Ohio 45433
AFAL (AVTA) R. D. Larson
Wright-Patterson AFB, Ohio 45433

AFAL (AVT) Dr H. V. Noble, Chief
Electronics Technology Division
Air Force Avionics Laboratory
Wright-Patterson AFB, Ohio 45433

Dr Robert E. Fontana
Head, Dept of Electrical Engineering
Air Force Institute of Technology
Wright Patterson AFB, Ohio 45433

Dept of Electrical Engineering
Clippinger Laboratory
Ohio University
Athens, Ohio 45701

Commanding Officer
Naval Avionics Facility
Indianapolis, Indiana 46241

Dr John C. Hancock, Head
School of Electrical Engineering
Purdue University
Lafayette, Ind 47907

Professor Joseph E. Rowe
Chairman, Dept of Electrical
Engineering
The University of Michigan
Ann Arbor, Michigan 48104

Dr G. J. Murphy
The Technological Institute
Northwestern University
Evanston, Ill 60201

Commanding Officer
Office of Naval Research Branch Office
219 South Dearborn St.
Chicago, Illinois 60604

Illinois Institute of Technology
Dept of Electrical Engineering
Chicago, Illinois 60616

Deputy for Res. and Eng (AMSE-DRE)
U. S. Army Weapons Command
Rock Island Arsenal
Rock Island, Illinois 61201

Commandant
U. S. Army Command & General Staff
College
Attn: Acquisitions, Library Division
Fort Leavenworth, Kansas 66027

Dept of Electrical Engineering
Rice University
Houston, Texas 77001

HQ AMD (AMR)
Brooks AFB, Texas 78235

USAFSAM (SMKOR)
Brooks AFB, Texas 78235

Mr B. R. Locke
Technical Adviser, Requirements
USAF Security Service
Kelly Air Force Base, Texas 78241

Director Electronics Research Center
The University of Texas at Austin
Eng-Science Bldg 110
Austin, Texas 78712

Department of Electrical Engineering
Texas Technological University
Lubbock, Texas 79409

Commandant
U. S. Army Air Defense School
Attn: Missile Sciences Div., C&S Dept
P. O. Box 9390
Fort Bliss, Texas 79916

Director
Aerospace Mechanics Sciences
Frank J. Seiler Research Laboratory (OAR)
USAF Academy
Colorado Springs, Colorado 80840

Director of Faculty Research
Department of the Air Force
U. S. Air Force Academy
Colorado Springs, Colorado 80840

Major Richard J. Gowen
Tenure Associate Professor
Dept of Electrical Engineering
U. S. Air Force Academy
Colorado Springs, Colorado 80840

Academy Library (DFSLB)
U. S. Air Force Academy
Colorado Springs, Colorado 80840

M. A. Rothenberg (STEPD-SC(S))
Scientific Director
Desert Test Center
Bldg. 100, Soldiers Circle
Fort Douglas, Utah 84113

Utah State University
Dept of Electrical Engineering
Logan, Utah 84321

School of Engineering Sciences
Arizona State University
Tempe, Ariz 85281

Commanding General
U. S. Army Strategic Communications Command
Attn: SCC-CG-SAE
Fort Huachuca, Arizona 85613

The University of Arizona
Dept of Electrical Engineering
Tucson, Arizona 85721

Cpt C. E. Baum
AFWL (WLRE)
Kirkland AFB, New Mexico 78117

Los Alamos Scientific Laboratory
Attn: Report Library
P. O. Box 1663
Los Alamos, N.M 87544

Commanding Officer
Atmospheric Sciences Laboratory
White Sands Missile Range, N. Mex 88002

Commanding Officer
(AMSEL-BL-WS-R)
Atmospheric Sciences Laboratory
White Sands Missile Range
New Mexico 88002

Chief, Missile Electronic Warfare Tech Area
(AMSEL-WL-M)
Electronic Warfare Laboratory, USACOM
White Sands Missile Range, N.M. 88002

Director
Electronics Sciences Lab
University of Southern California
Los Angeles, Calif 90007

Engineering & Mathematical Sciences Library
University of Californaia at Los Angeles
405 Hilgred Avenue
Los Angeles, Calif. 90024

Aerospace Corporation
P. O. Box 95085
Los Angeles, California 90045
Attn: Library Acquisitions Group

Hq SAMSO (SMTTA/Lt Belate)
AF Unit Post Office
Los Angeles, Calif. 90045

Dr Sheldon J. Wells
Electronic Properties Information Center
Mail Station E-175
Hughes Aircraft Company
Culver City, California 90230

Director, USAF PROJECT RAND
Via: Air Force Liaison Office
The RAND Corporation
Attn: Library D
1700 Main Street
Santa Monica, California 90406

Deputy Director and Chief Scientist
Office of Naval Research Branch Office
1030 East Green Street
Pasadena, California 91101

Aeronautics Library
Graduate Aeronautical Laboratories
California Institute of Technology
1201 E. California Blvd.
Pasadena, California 91109

Professor Nicholas George
California Inst. of Technology
Pasadena, California 91109

Commanding Officer
Naval Weapons Center
Corona Laboratories
Attn: Library
Corona, California 91720

Dr F. R. Charvat
Union Carbide Corporation
Materials Systems Div.
Crystal Products Dept.
8888 Balboa Avenue
P.O. Box 23017
San Diego, Calif 92123

Hollander Associates
P. O. Box 2276
Fullerton, California 92633

Commander
U. S. Naval Missile Center (56322)
Point Mugu, California 93041

W. A. Eberspacher, Associate Head
Systems Integration Division
Code 5340A, Box 15
U. S. Naval Missile Center
Point Mugu, California 93041

Sciences-Engineering Library
University of California
Santa Barbara, California 93106

Commander (Code 753)
Naval Weapons Center
Attn: Technical Library
China Lake, California 93555

Library (Code 2124)
Technical Report Section
Naval Postgraduate School
Monterey, California 93940

Glen A. Myers (Code 52Mv)
Assoc. Professor of Elec. Engineering
Naval Postgraduate School
Monterey, California 93940

Dr Leo Young
Stanford Research Institute
Menlo Park, Calif. 94025

Lenkurt Electric Co., Inc.
1105 County Road
San Carlos, California 94070
Attn: Mr E. K. Peterson

Director
Microwave Laboratory
Stanford University
Stanford, California 94305

Director
Stanford Electronics Laboratory
Stanford University
Stanford, California 94305

Director, Electronics Research Laboratory
University of California
Berkeley, California 94720

DELETE:

LTC H.W. Jackson

REPLACE WITH:

Dr. L. A. Wood, Director
Electronic & Solid State Sciences
Air Force Office of Scientific Res.
1400 Wilson Boulevard
Arlington, Virginia 22209

DELETE:

Mr. Norman J. Field

REPLACE WITH:

Mr. Norman J. Field, AMCPM-AA-PM
Chief, Program Management Division
Project AACOMS, USAECOM, Bldg. 2525
Fort Monmouth, New Jersey 07703

DELETE:

Mr. Robert O. Parker, AMSEL-RD-S

REPLACE WITH:

Mr. I.A. Balton, AMSEL-XL-D
Executive Secretary, TAC/JSEP
U.S. Army Electronics Command
Fort Monmouth, New Jersey 07703

ADDENDUM:

Dr. H.K. Ziegler, Chief Scientist
Army Member TAC/JSEP (AMSEL-SC)
U.S. Army Electronics Command
Fort Monmouth, New Jersey 07703

Dr. Billy Welch
USAFSAM (SMC)
Brooks AFB, Texas 78235

DELETE:

Director, U.S. Army Material
Concepts Agency
Washington, D.C. 20315

REPLACE WITH:

Director
USA Advanced Materiel Concepts
Agency
2461 Eisenhower Avenue
Alexandria, Va. 22314

Security Classification

## DOCUMENT CONTROL DATA - R & D

*(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)*

| | |
|---|---|
| 1. ORIGINATING ACTIVITY *(Corporate author)*<br>University of Illinois<br>Coordinated Science Laboratory<br>Urbana, Illinois 61801 | 2a. REPORT SECURITY CLASSIFICATION<br><br>2b. GROUP |

3. REPORT TITLE

A DATA STRUCTURE FOR COGNITIVE INFORMATION RETRIEVAL

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*

K.O. Biss, R.T. Chien, F.A. Stahl

| 6. REPORT DATE | 7a. TOTAL NO. OF PAGES | 7b. NO. OF REFS |
|---|---|---|
| October 1970 | 17 | 12 |

| | |
|---|---|
| 8a. CONTRACT OR GRANT NO.<br>DAAB 07-67-C-0199; also in part<br>b. PROJECT NO. OE 1-7-071213-4557<br>c.<br>d. | 9a. ORIGINATOR'S REPORT NUMBER(S)<br>R - 493<br>9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*<br>UILU-ENG 70-238 |

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

| 11. SUPPLEMENTARY NOTES | 12. SPONSORING MILITARY ACTIVITY |
|---|---|
| | Joint Services Electronics Program<br>thru U.S. Army Electronics Command<br>Fort Monmouth, New Jersey 07703 |

13. ABSTRACT

A new data structure developed in connection with a natural language question-answering system is described. The data structure is based upon a new high-order calculus. It allows a great degree of expressiveness and is suitable for extensive logical deduction.

The inadequacies of low-order schemes for the representation of natural language information are given. A formal description of the new high-order structure which overcomes these inadequacies is then presented along with the properties that make it more suitable and attractive for machine processing of natural language information.

DD FORM 1 NOV 65 1473

Security Classification

| 14. KEY WORDS | LINK A | | LINK B | | LINK C | |
|---|---|---|---|---|---|---|
| | ROLE | WT | ROLE | WT | ROLE | WT |
| Information Retrieval | | | | | | |
| Natural Language | | | | | | |
| Question-Answering | | | | | | |
| Data Structuring | | | | | | |
| Theorem Proving | | | | | | |
| Deduction | | | | | | |