Automatica, Vol. 19, No. 6, pp. 759–766, 1983
Printed in Great Britain.

0005-1098/83 $3.00 + 0.00
Pergamon Press Ltd.

Brief Paper

An Expert Intermediary System for Interactive Document Retrieval*

GIOVANNI GUIDA†‡ and CARLO TASSO†§

Key Words—Artificial intelligence; man–machine systems; online databases; natural language dialogue; reasoning; expert systems; information retrieval; document retrieval.

Abstract—Constructing natural language interfaces to computer systems often requires achievement of advanced reasoning and expert capabilities in addition to basic natural language understanding. In this paper the above issue is tackled in the frame of an actual application concerning the design of a natural language interface for interactive document retrieval. After a short discussion of the peculiarities of this application, which requires both natural language understanding and reasoning capabilities, the general architecture and fundamental design criteria of a system presently being developed at the University of Udine are presented. The system, named IR–NLI, is aimed at allowing non-technical users to directly access through natural language the services offered by online databases. Attention is later focused on the basic functions of IR–NLI, namely understanding, dialogue and reasoning. An example of interaction with IR–NLI is fully worked out to introduce the main features of the system. Knowledge representation methods and algorithms adopted are then illustrated. Perspectives and direction for future research are also discussed.

1. *Introduction*

NATURAL language interfaces and expert systems constitute two of the most promising applications of artificial intelligence. Their technologies are often combined together in the design of complex systems, where natural language is used to provide easy access to consultation services for casual users (Shapiro and Kwasny, 1975). Still the issue of conceiving in an unitary way understanding and deduction capabilities is quite new and shows several interesting facets. It would be very desirable to model in an unitary way these activities, which appear strongly connected in humans (Smith, 1980). In fact, understanding is often aimed at acquiring knowledge to be utilized in deductive and inference activities, while, on the other hand, reasoning on domain-specific and commonsense knowledge is required in several cases to carry out the comprehension process. From a more technical point of view, it is argued that a common knowledge representation for both understanding and reasoning functions could be appropriate to ensure a good level of integration and performance. Moreover, a unitary internal representation could suggest the design of homogeneous, mainly rule-based, algorithms for the basic processing steps of these activities.

This paper is concerned with one topic which belongs partly to the area of natural language processing and partly to that of knowledge-based consultation systems: namely, natural language reasoning. This is intended as a basic capability in natural language communication, which is aimed at capturing speaker's goals and intentions, often lying behind the mere literal meaning of the utterance.

In this work we explore the main implications of natural language reasoning in the frame of an actual application which fits quite well the combined issues of language comprehension and expertise delivery: the natural language access to interactive document retrieval systems (Waterman, 1978; Pollitt, 1981; Marcus, 1981; Anderson, 1977). In fact, it is well known that this task requires not only a basic natural language understanding capability, but also the ability of performing the intermediary's role: analysing user's requests, reasoning on them, capturing user's needs and wants, and generating an appropriate search strategy.

In particular, we shall present the detailed design of a system, named IR–NLI (Information Retrieval–Natural Language Interface), which is being developed at the University of Udine, and we shall discuss its main original features. The topic of natural language reasoning is first shortly illustrated from a conceptual point of view and compared to related proposals. The main features of the chosen application domain are then described, and the specifications of IR–NLI are stated. We later turn to the architecture of the system and we go further into a detailed account of the structure of its knowledge bases and mode of operation. Particular attention is devoted to the two fundamental modules: REASONING and UNDERSTANDING AND DIALOGUE. A sample search session with IR–NLI is fully worked out to show the system in operation. A critical evaluation of the work done is then presented, and main lines of the future development of the project are outlined.

2. *The theoretical frame*

Reasoning, intended as the generic capability of making inferences, is not a new issue in the area of natural language processing. Anaphora resolution, indirect speech acts analysis, ellipsis expliciting, for example, all require inference capabilities.

By 'reasoning' we denote here a very specific function involved in natural language man–machine communication. We assume that, when a person communicates with a machine always there is some *goal* behind speech acts, which moves him to communicate and determines the utterances he uses in the communication—for example, to obtain some desired behaviour or service from the machine. Such goal is not always clear, precise, and univocally expressed in the man–machine dialogue. In fact, the user is often only poorly acquainted with the functions and use of the machine, or, also, not very sure of his own needs and wants. Furthermore, even if the goal of the communication is definitely clear to the user, the way he expresses it to the machine is often incomplete, or ambiguous, or partial since a lot of common knowledge is implicitly understood, or hardly understandable due to the use of a different terminology. If we agree that comprehension requires capturing or reconstructing the goal of the speaker (Grosz, 1979; Allen and Perrault, 1980), it is straightforward that a literal understanding of natural language expressions is only a small part of the entire comprehension task.

---

* Received 14 January 1983; revised 24 March 1983; revised 21 June 1983; revised 23 August 1983. The original version of this paper was presented at the IFAC/IFIP/IFORS/IEA Conference on Analysis, Design, and Evaluation of Man–Machine Systems which was held in Baden-Baden, F.R.G. during September 1982. The published proceedings of this IFAC meeting may be ordered from Pergamon Press Ltd, Headington Hill Hall, Oxford OX3 0BW, U.K. This paper was recommended for publication in revised form by editor A. Sage.

† Istituto di Matematica, Informatica e Sistemistica Università di Udine, Udine, Italy.

‡ Also with Milan Polytechnic Artificial Intelligence Project, Milano, Italy.

§ Also with CISM, International Center for Mechanical Sciences, Udine, Italy.

Therefore, we distinguish in the natural language comprehension activity two phases: a surface comprehension which only aims at representing the literal content of a natural language expression into a formal internal representation, and a deep comprehension which moves beyond surface meaning to capture the goals and intentions which lie behind the utterance. The process that brings from surface to deep comprehension is just what we name here *reasoning* activity. Differently from Winograd (1980), reasoning is not, in our model, something that takes place after understanding is completed and aims at developing deductions on facts and concepts acquired. Reasoning is a basic part of comprehension and involves not only linguistic capabilities (understanding and dialogue) but also inference, analogy, generalization, expertise, etc., on common sense and domain specific knowledge. Figure 1 shows a graphic representation of the basic relationships between understanding and reasoning, and illustrates how reasoning moves the internal representation of an utterance from a first point, corresponding to surface comprehension, to a second one which represents deep comprehension.

3. *The application domain*

In this section we present an application domain where the topic of natural language reasoning plays a fundamental role, namely, natural language access to online information retrieval services (Lancaster, 1979; Meadow and Cochrane, 1981). Online services allow interested users to solve information problems by selecting and retrieving relevant documents stored in very large bibliographic or factual databases. Generally, end-users are unwilling (or unable) to search personally, and directly access these large files, but they rely on the assistance of a skilful information professional: the intermediary who knows how to select appropriate databases and how to design good search strategies for the retrieval of the desired information, and how to implement them in a suitable formal query language. Usually, the interaction between end-user and intermediary begins with a presearch interview aimed at precisely clarifying the content and the objectives of the information need. On the basis of the information gathered, the intermediary chooses the most suitable data bases and, with the help of searching referral aids, such as thesauri, directories, etc., he devises the search strategy to be executed by the information retrieval system. The output of the search is then evaluated by the end-user, who may propose a refinement and an iteration of the search.

Going deeper in analysing the peculiarities of this application, let us contrast online informaton retrieval to the enquiry of a database management system (Waltz, 1978; Landauer and coworkers, 1982). We consider two main points. First, the information stored in a database has a definite logical structure which supports the access to data and helps the user in finding desired information items. On the other hand, in the information retrieval case, the stored records (which contain: title, authors, abstract, key words, etc. of the documents) identify the content of the documents only in a partial and unstructured way. This implies that retrieval of requested information is much more difficult, since it relies on several loosely defined factors, such as domain specific knowledge, knowledge about indexing criteria, availability of updated and complete searching referral aids, working experience on the particular database, etc. This first feature is clearly captured by the two traditional criteria of precision and recall (Meadow and Cochrane, 1981), which are used to evaluate the quality of a search strategy. Second, while a user of a database generally looks for precise information to be extracted from the stored files, a user of an online service desires to get information on a given topic which does not immediately correspond to some stored object; what records to extract in order to match the user's request is usually a non-trivial problem, that has to be solved by the intermediary.

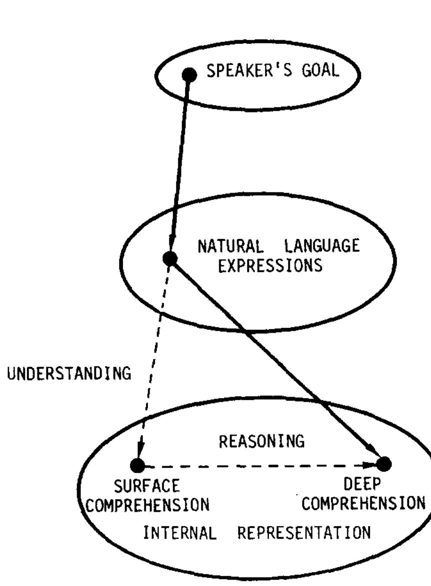

FIG. 1. Understanding and reasoning in natural language comprehension.

We claim that the intermediary's task represents a good example of the issues of natural language reasoning, particularly for what concerns the ability of understanding language user's requests and of reasoning on their linguistic and semantic facets in order to fully capture user's needs and goals. Besides, it has to be stressed that the intermediary should also possess other important skills, including expertise and precise knowledge about database content and organization, about system query languages and access procedures, and about how to plot and construct an adequate search strategy. The above illustrated characteristics motivate the design of a natural language expert system for interfacing online document retrieval systems.

4. *System specifications and architecture*

IR–NLI is conceived as an interactive interface to interactive document retrieval systems, supporting English language interaction. It should be able to manage a dialogue with the user on his information needs and to construct an appropriate search strategy. More precisely, IR–NLI is aimed at meeting the needs of non-technical users who are not acquainted with online searching. For this purpose, three different capabilities are requested. First, the system has to be an expert of online searching, i.e. it must be knowledgeable of the intermediary's professional skill. Second, it must be capable of understanding natural language and of carrying out a dialogue with the user. Third, it has to be capable of reasoning on specific terminology in order to capture the information needs of the user and to formulate them with appropriate terms in a given formal query language.

Furthermore, among the most application-oriented goals of IR–NLI are the following capabilities:

to operate as a front-end to several available data bases and choose among them the most appropriate to answer the user's requests;

to deal with several specific subject domains;

to utilize the different query languages needed to access the available databases.

The general architecture of IR–NLI is shown in Fig. 2. The kernel of the system is constituted by the UNDERSTANDING AND DIALOGUE and REASONING modules, which are connected to the four knowledge bases available to the system.

The *UNDERSTANDING AND DIALOGUE* module is devoted to perform activities mostly of linguistic concern. First, it translates the natural language user's requests into a formal *internal representation* (IR). Second, it manages (under the control of the REASONING module) the dialogue with the user by generating appropriate queries and by translating his replies, thus expanding the IR with new information. The UNDERSTANDING AND DIALOGUE module utilizes for its operation a *vocabulary* (VOC) and a base of *linguistic rules* (LR).

The *REASONING* module is devoted to devise the top-level choices concerning the intelligent operation of the system and to control their execution. It utilizes for its activity a base of *expert rules* (ER) which concern the evaluation of user's requests, the management of the presearch interview, the expansion of the IR, the selection of a suitable approach for generation of the search strategy, the activation of the dialogue with the user, and the collection of the information needed to formulate a suitable search strategy. The subject knowledge necessary to all these tasks is contained in a base of *domain specific knowledge* (DSK). The *FORMALIZER* module, after the REASONING module has completed its activity, constructs from the fully expanded IR the output *search strategy* to be executed for accessing the online database. The FORMALIZER utilizes, for its operation, knowledge about the formal language needed to interrogate the online data base and operates through a simple syntax-directed schema. It is conceived as a parametric translator capable of producing search strategies in several languages for accessing online services, such as SDC ORBIT, Euronet DIANE, Lockheed DIALOG, etc.

5. *IR–NLI in operation: an example*

Before going into more details concerning the mode of operation and the precise tasks of each module of IR–NLI, we present in this section a sample search session. We will focus more specifically on the internal representation, on the dialogue with the user, and on the basic reasoning mechanisms.

The example refers to the domain of computer science, and starts with the submission by the user of the following natural language request:

user > I would like to have references about computer applications in warehousing. I am interested only in English documents, not older than three years.

The UNDERSTANDING AND DIALOGUE module analyses the request and extracts from it the information which describes the search topic. The first version of the IR is shown in Fig. 3. The REASONING module initializes then the presearch interview, aimed at gathering from the user further information to be utilized in constructing the search strategy. The mode of operation in which the REASONING module manages this activity is basically data-driven; in fact, the interaction with the user is fired by the detection of empty (or only partially filled in) slots in the IR. For this task a subset of the ER is utilized, namely, the presearch interview rules (PIR).

In our example the system asks the user a few questions and expands the current IR:

IR–NLI > specify the scope of the search:
comprehensive: retrieval of all relevant items
limited: retrieval of only relevant items
sample: retrieval of some relevant items
user > comprehensive
('high recall' is inserted in the OBJECTIVES subframe of the IR)

CONCEPTS :	
C.1	computer applications
C.2	warehousing
ASPECTS :	
LIMITATIONS :	
L.1	language = English
L.2	year ≥ 1980
SEARCH MODE :	
OUTPUT :	
OBJECTIVES :	
SELECTED APPROACH :	
PLANNED TASKS :	

FIG. 3. IR of the sample search session (1).

IR–NLI > specific treatment desired (general, practical, theoretical, ...)
user > practical
('treatment = practical' is stored in the LIMITATIONS subframe L.3)
IR–NLI > output desired:
online/offline
format: bibliographic references with/without abstract, full format, ...
sorted
maximum number of items desired
user > full format, offline, not sorted, at most 50 postings
('full format', 'offline', 'not sorted', and 'max = 50' are filled in the OUTPUT subframe)
IR–NLI > list key terms, synonyms, and spelling variations for each concept
C.1: computer applications

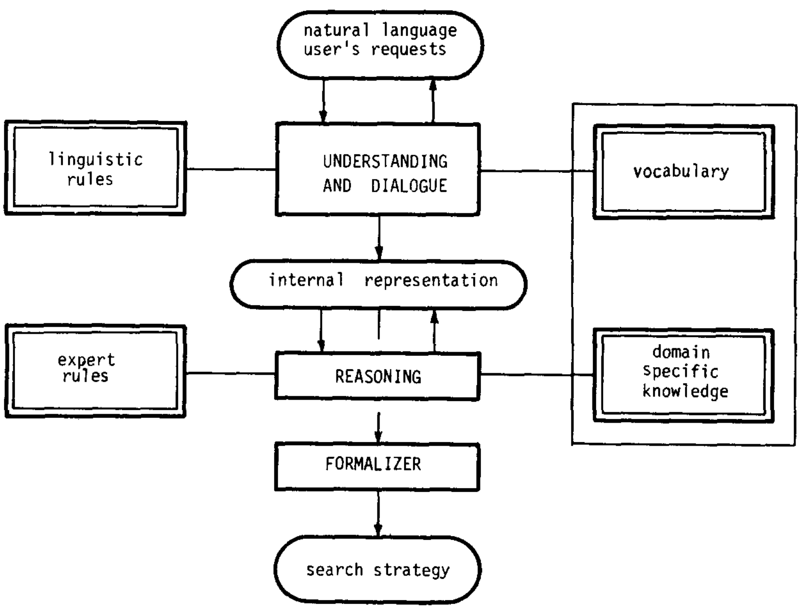

FIG. 2. Overall architecture of IR–NLI.

user > management information systems
IR–NLI > C.2: warehousing
user > none
('management information systems' is inserted in the KEY TERMS slot of the subframe C.1)
IR–NLI > supply any particular point of view or specific aspect to take into account (design, performance, reliability, security, ...)
user > cost/benefit evaluation
('cost/benefit evaluation' is stored in the ASPECTS subframe A.1)
IR–NLI > for each concept and aspect specify the degree of relevance to the search (score from 1 to 10)
C.1: computer applications
user > 10
IR–NLI > C.2: warehousing
user > 10
IR–NLI > A.1: cost/benefit evaluation
user > 5
(the values 'DI = 10', 'DI = 10', and 'DI = 5' are inserted in the subframes of C.1, C.2, and A.1, respectively, to denote the degree of interest of the user in each specific facet of the search).

The presearch interview is now closed and the IR is completed with information extracted from the DSK, including posting count (PC), generality level (GL), and controlled term flag (CT), of each concept and aspect, which will be used later for selecting the search approach. The current IR at this step is shown in Fig. 4.

The selection of the search approach is made by means of high-level rules (HLR), a subset of the ER. In this case the following rule applies:

IF (posting count of all concepts *high*) AND (generality level of all concepts *similar*)
THEN select building block approach

where *high* means greater than a given integer *m*, depending on the specific subject domain; and *similar* denotes that the maximum difference between any two levels must not exceed a given integer *n*.

This choice is then inserted in the SELECTED APPROACH subframe of the IR and taken into consideration for matching with medium-level rules (MLR). These are responsible for further enlarging and refining terminology by accessing the DSK network. Several MLR are fired in our example; we show here just a few of them, for example:

IF (building block) AND (high recall) AND (number of terms in concept C < max)
THEN *expand* C

where max is an integer depending on the PC and DI of the concept C;

IF (building block) AND (high recall)
THEN *search free text*

'*expand* C.1' and '*expand* C.2' are inserted in the PLANNED TASKS subframe and *search free text* in the SEARCH MODE subframe of the IR. The first of these two subframes is utilized to fire low-level rules (LLR): these are concerned with the specific actions to be taken in accessing the DSK. In more detail, these actions are of two kinds:

- move inside the DSK, from one node towards other nodes connected by specific kinds of arcs;
- insert in the IR the new terminology acquired during this navigation, possibly after a validation of the user.

In order to illustrate this point, consider Fig. 5, which shows a small portion of the DSK, centered around the first concept (C.1) of our example. When the planned task '*expand* C.1' is executed, the following LLR, among others, is fired

IF *expand* C
THEN *parallel* C.

CONCEPTS :				
C.1	computer applications			
	DI = 10	GL = 8	PC = 12,005	CT = true
	KEY-TERMS : management information systems			
C.2	warehousing			
	DI = 10	GL = 6	PC = 2,734	CT = true
ASPECTS :				
A.1	cost/benefit evaluation			
	DI = 5			CT = false
LIMITATIONS :				
L.1	language =English			
L.2	year $\geq$ 1980			
L.3	treatment = practical			
SEARCH MODE :				
OUTPUT :				
full format	offline	not sorted	max = 50	
OBJECTIVES : high recall				
SELECTED APPROACH :				
PLANNED TASKS :				

FIG. 4. IR of the sample search session (2).

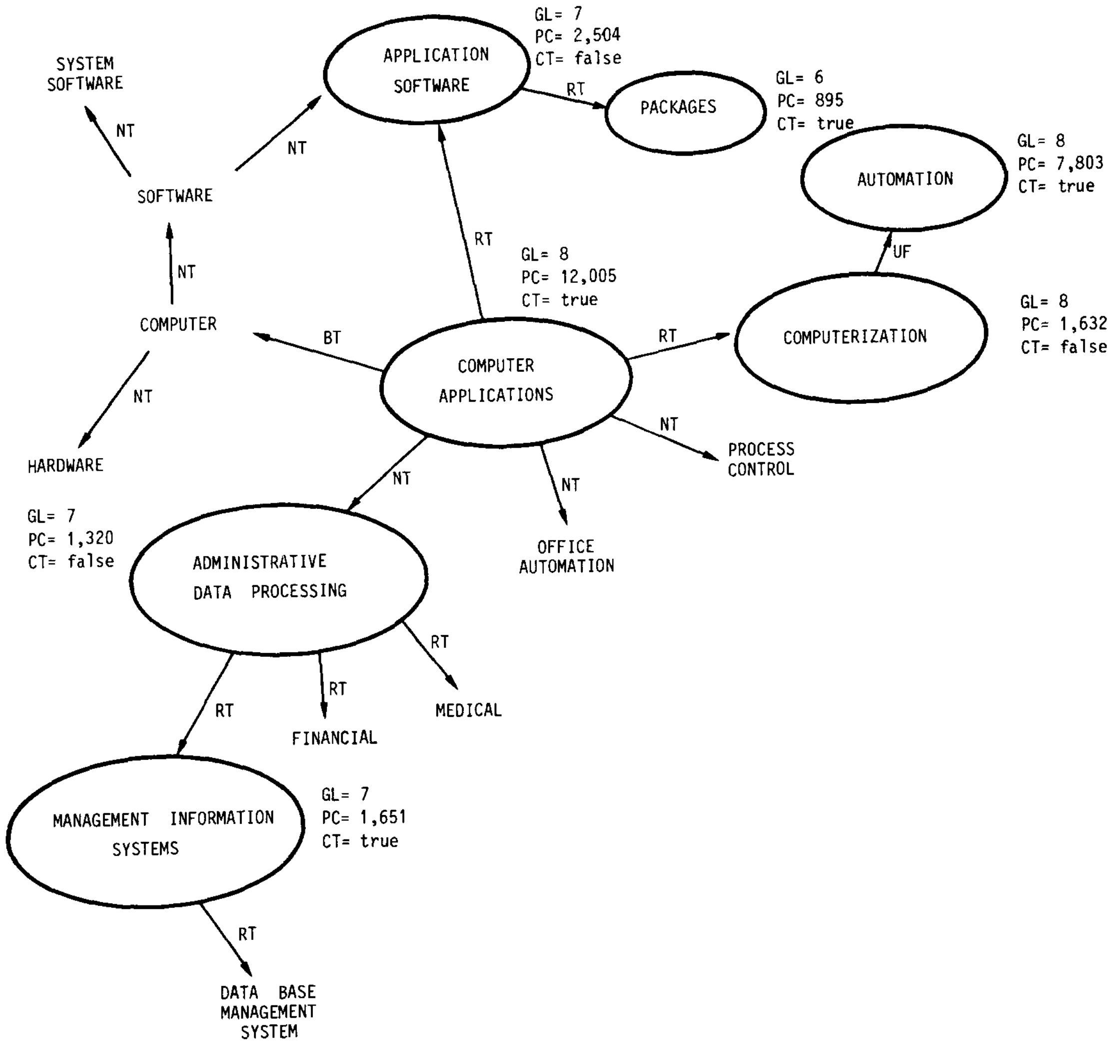

FIG. 5. DSK (partial) around the term 'computer applications'.

The command '*parallel* C.1' is now inserted in the PLANNED TASKS subframe of the IR ('*expand* C.1' is removed from the IR, once all appropriate LLR have been fired), and then it is executed through a fixed (parametric) procedure. In our case, it moves in the DSK along RT (related term) and UF (used for term) arcs, and captures the new terminology 'application software', 'packages', 'computerization', and 'automation', which is added to the IR in the subframe of C.1.

Figures 5 and 6 represent a subset of the DSK around the concepts C.1 and C.2, and show the terminology reached by the LLR during the execution of '*expand* C.1' and '*expand* C.2'.

Among the LLR which are activated when the expansion of the IR is almost completed, we mention also the following one

IF concept or aspect C is *fully expanded*
THEN *compile* C.

When this rule is fired, the commands *compile* C.1, *compile* C.2, and *compile* A.1 are inserted in the IR and then executed. Their action refers to further refinement of the terminology, including respelling terms, truncating words which terminate by suffixes such as '-ation', '-ing' and so on, respacing compound words, etc.

When no more ER are fireable, the expansion of the IR is completed. The final version of the IR in our example is shown in Fig. 7.

The knowledge contained in the final IR is used by the FORMALIZER to construct the output search strategy. The search strategy obtained in our example is shown in Fig. 8 (a dialect of EUROLANGUAGE has been used).

6. *Knowledge representation and basic algorithms*

In this section we describe the general aspects concerning the overall mode of operation of IR–NLI. As a background, we first analyse the main features and organization of the knowledge bases, namely, the vocabulary (VOC) and the domain specific knowledge (DSK).

Let us begin with the DSK. The purpose of this knowledge base is to store information about the domain covered by the online data base to which IR–NLI refers. This information presents two aspects: a semantic facet concerning what concepts are in the data base and how they relate to each other, and a linguistic one concerning how the concepts are currently expressed through appropriate terms. The internal structure of the DSK reflects and generalizes to some extent that of classical searching referral aids (in particular, thesauri and subject classifications). At a logical level, it is constituted by a labelled directed network in which *nodes* represent concepts and directed *arcs* represent relations between concepts. Each node contains:

a *term;*

a Boolean flag denoting whether the term is controlled or not (*CT*);

a field that stores the posting count (*PC*), i.e. the number of items of the database in which the term appears;

a generality level (*GL*), which represents the degree of specificity of the term in a hierarchical subject classification.

Arcs generally denote the usual cross-reference relationships utilized for structuring thesauri; these include: *BT* (broader term), *NT* (narrower term), *RT* (related term), *UF* (used for).

This structure is conceived to be directly obtained (possibly in

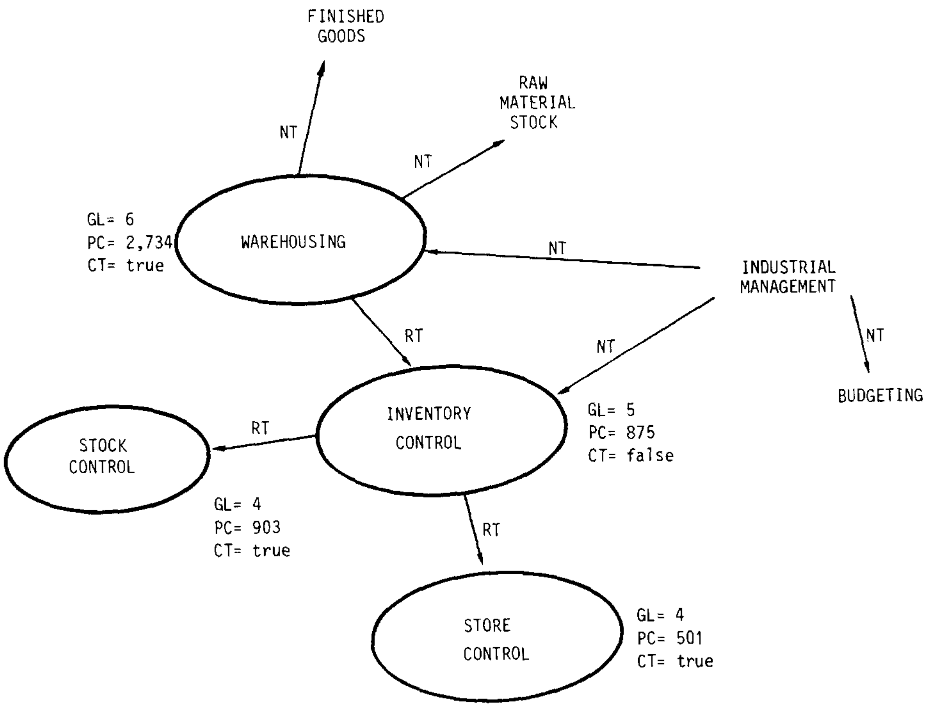

FIG. 6. DSK (partial) around the term 'warehousing'.

CONCEPTS :				
C.1	computer.applicat*	applicat*.software		
	computeriz*	administrative.data.process*		
	automat*	package*		
	KEY-TERMS : management.information.system*			
	DI = 10	GL = 8	PC = 12,005	CT = true
C.2	warehous*	inventory.control	stock*.control	
	stor*.control			
	DI = 10	GL = 6	PC = 2,734	CT = true
ASPECTS :				
A.1	cost*.benefit*.evaluat*			
	DI = 5			CT = false
LIMITATIONS :				
L.1	language = English			
L.2	year ≥ 1980			
L.3	treatment = practical			
SEARCH MODE : free text				
OUTPUT :				
full format	offline	not sorted	max = 50	
OBJECTIVES : high recall				
SELECTED APPROACH :				
PLANNED TASKS :				

FIG. 7. IR of the sample search session (3).

```
1.  FIND   COMPUTER.APPLICAT *
2.  FIND   APPLICAT*SOFTWARE
3.  FIND   COMPUTERIZ *
4.  FIND   ADMINISTRATIVE.DATA.PROCESS *
5.  FIND   AUTOMAT *
6.  FIND   PACKAGE *
7.  FIND   MANAGEMENT.INFORMATION.SYSTEM *
8.  FIND   S=1 OR S=2 OR S=3 OR S=4 OR S=5 OR S=6 OR S=7
9.  FIND   WAREHOUS *
10. FIND   INVENTORY.CONTROL
11  FIND   STOCK*CONTROL
12  FIND   STOR*.CONTROL
13  FIND   S=9 OR S=10 OR S=11 OR S=12
14  FIND   COST*.BENEFIT*.EVALUAT*
15  FIND   S=8 AND S=13 AND S=14
16  LIMIT  S=15 / LANG = ENGLISH
17  LIMIT  S=16 / TREAT = PRACTICAL
18  LIMIT  S=17 / YEAR ≥ 1980
19  PRINT  F1, 1...50
```

FIG. 8. Search strategy obtained in the sample search session.

a partially automatic way through appropriate data conversion programs) from available searching referral aids and online thesauri.

We turn now to the VOC. This knowledge base is aimed at supplying all information concerning natural language that is needed to understand user's requests. According to the mode of operation of the UNDERSTANDING AND DIALOGUE module (see later), it contains the lexicon of the application domain which is currently considered. Each record of the lexicon contains a *word* of the language, its *semantic type* (concept, connective, function), and its *meaning*. The semantic type denotes the role of a word in a sentence; namely

denoting a term of the data base (concept);

defining a particular relation between different concepts in user's requests (connective);

specifying a particular function that the user desires to obtain from the information retrieval system (function).

The meaning of a word is expressed as a pointer to a term of the DSK in the case of a word of type concept, as a flag for the activation of linguistic rules in the case of a connective or a function.

Let us note that, in order to avoid useless duplication of information in the DSK and VOC, a shared directory of entry words is utilized for both knowledge bases.

We are now able to illustrate in some detail the general mode of operation of IR–NLI.

The task of the REASONING module can be considered from two different points of view;

an external one, which concerns performing the intermediary's activity;

an internal one, which relates to the management and control of the inference process and of the UNDERSTANDING AND DIALOGUE module.

On the basis of these specifications, it must contain expert capabilities and behave as a consultation system for information retrieval (Pollitt, 1981). The basic mode of operation of this module is organized around the following three main steps, which reflect the usual practice of online information searching (Lancaster, 1979; Meadow and Cochrane, 1981)

1. Perform presearch interview.
2. Select search approach.
3. Implement search approach and expand IR.

The internal representation (IR) adopted is unique throughout the operation of the system and it is consistuted by a frame, initially filled in by the UNDERSTANDING AND DIALOGUE module, then further refined and expanded by the REASONING module, and finally utilized by the FORMALIZER to generate the output search strategy. This frame is structured into subframes in such a way to contain, classified under different headings, any information that is relevant for searching an online data base, and to allow an effective pattern-matching for the selection of the appropriate expert rules (ER) to be fired. More specifically, the IR contains the following main subframes:

CONCEPTS
  KEY-TERMS
  DI (degree of interest)
  GL (generality level)
  PC (posting count)
  CT (controlled term flag)
ASPECTS
  DI (degree of interest)
  GL (generality level)
  PC (posting count)
  CT (controlled term flag)
LIMITATIONS
  language
  year
  author
  .
  .
  .
  treatment
SEARCH MODE
OUTPUT
OBJECTIVES
SELECTED APPROACH
PLANNED TASKS.

The first five subframes are devoted to store the information necessary to construct the search strategy, while the last three contain information used for only the operation of the REASONING module. More precisely, the last two constitute the working memory (agenda) which contains the intermediary results of the activation of the ER.

The operation of the REASONING module is basically pattern-directed; namely, the particular activities to be performed and the way in which the UNDERSTANDING AND DIALOGUE module is activated are determined by the content of the current IR (or of some subframes of it) which it matched with an appropriate subset of the ER.

For classes of ER are provided:

presearch interview rules (PIR)

high-level rules (HLR)

medium-level rules (MLR)

low-level rules (LLR)

which apply to the three basic phases of operation of the REASONING module introduced above (the last two classes both refer to phase 3). The activity of the REASONING module can now be represented in a more detailed way through the following abstract program:

module REASONING
  activate UNDERSTANDING AND DIALOGUE
  ⟨generation of IR from initial user's requests⟩
  perform presearch interview
  ⟨analysis of the current IR and selection through PIR of subframes which could be appropriately filled in with new information⟩
  activate UNDERSTANDING AND DIALOGUE
  ⟨engagement of a suitable dialogue with the user for gathering additional information about search content and objectives: concepts, aspects, keyterms, limitations, scope, treatment, output, etc.⟩
  ⟨access to DSK⟩
  ⟨expansion of IR⟩
  select search approach
  ⟨selection through HLR of the approach which best fits the current IR⟩
  implement search approach
  ⟨activation and firing of the appropriate MLR and LLR which match the current IR⟩
  ⟨access to DSK⟩
  activate UNDERSTANDING AND DIALOGUE

<validation of the currently expanded IR>
<expansion of IR>
activate FORMALIZER
<generation of the search strategy from fully expanded IR>
endmodule.

We turn now to the UNDERSTANDING AND DIALOGUE module. The purpose of this module is twofold: to translate user's requests into the IR; and to generate queries to the user and to understand his answers, i.e. to manage a bounded-scope dialogue.

The conception of this module strongly relies, for what concerns the understanding function, on the experience previously developed by the authors with the NLI project, and is organized around the concept of semantics-directed and goal-oriented parsing (Guida and Tasso, 1982a). Its mode of operation is mainly rule-based: a main parsing algorithm performs the most elementary steps of the analysis (search in the vocabulary, construction of a basic tentative internal representation, validation of the basic internal representation) and manages a pattern-directed invocation of LR for resolution of critical events (e.g. ambiguity, ellipsis, anaphoric reference, indirect speech, etc.). An important feature of the understanding function is the ability to solve critical situations by engaging the user in a clarification dialogue activated by some of the above mentioned LR, to gather additional information which is necessary to correctly understand the input natural language requests.

For what concerns the dialogue function, it relies on two strictly connected activities: generation of a query, according to some request from the REASONING module, through assembly and completing of available parametric text fragments; and understanding of the user's answer and refinement (i.e. validation, updating or completing) of the current IR.

Let us stress that, according to the basic goal-oriented conception of the parsing mechanism of the UNDERSTANDING AND DIALOGUE module, the understanding activity performed in the frame of the dialogue function is strongly directed by knowledge of the query that the system has asked the user and, therefore, of expected information to be captured in the answer.

7. *Conclusion*

In the paper the main features of the IR–NLI system have been presented. The project is now entering an experimental phase, which will be carried on a VAX 11/780 system in Franz Lisp.

The design activity so far developed (Guida and Tasso, 1982b, 1983) has reached a quite assessed point, so that future work on this topic will be mainly concerned with removal of the restrictions and working hypotheses considered in the current first phase and with refinement of implementation details. We also plan to develop in the near future a complete prototype version of the system to be connected to a real online system in the frame of a strictly application-oriented interest.

The research activity will be focused, on the other hand, on several issues that deserve further investigation. Among these we mention:

the development of more flexible and robust dialogue capabilities, including limited justification of the mode of operation of the system (Webber and Joshi, 1982);

the enrichment of the ER base to cover all most commonly used approaches (Lancaster, 1979) and tactics (Bates, 1979) for online information retrieval;

the design of appropriate meta-rules to be used for a more flexible and effective management of the ER base;

the design of a learning subsystem connected to the REASONING module, devoted to keep track of previous search sessions and to analogize from experience in devising and implementing a search approach.

References

Allen, J. F. and C. R. Perrault (1980). Analysing intention in utterances. *Artificial Intelligence*, **15**.

Anderson, R. H. (1977). The use of production systems in RITA to construct personal computer 'agents'. *ACM SIGART Newsletter*, **63**, 23.

Bates, M. J. (1979). Information search tactics, *J. Am. Soc. Inf. Sci.*, **30**, 205.

Grosz, B. D. (1979). Utterance and objective: issues in natural language communication. *Proceedings 6th International Joint Conference on Artificial Intelligence*, Tokyo, pp. 1067–1076.

Guida, G. and C. Tasso (1982a). NLI: a robust interface for natural language person–machine communication. *Int. J. Man–Machine Studies*, **17**, 417.

Guida, G. and C. Tasso (1982b). Natural language access to online data bases. In *Cybernetics and Systems Research, Proceedings 6th European Meeting on Cybernetics and System Research*, Wien, R. Trappl (Ed.). North-Holland, Amsterdam, pp. 891–896.

Guida, G. and C. Tasso (1983). IR–NLI: an expert natural language interface to online data bases. *Proceedings ACL Conference on Applied Natural Language Processing*, Santa Monica, CA, pp. 31–38.

Lancaster, F. W. (1979). *Information Retrieval Systems.* John Wiley, New York.

Landauer, T. K., S. T. Dumais, L. M. Gomez and G. W. Furnas (1982). Human factors in data access. *Bell System Technical J.*, **61**.

Marcus, R. S. (1981). An automated expert assistant for information retrieval in the information community: An alliance for progress. *Proceedings 44th A.S.I.S. Annual Meeting*, Vol. 18, pp. 270–273.

Meadow, C. T. and P. A. Cochrane (1981). *Basics of Online Searching.* John Wiley, New York.

Pollitt, A. S. (1981). An expert system as an online search intermediary. In *Proceedings 5th International Online Information Meeting*, London, pp. 25–32.

Shapiro, S. C. and S. C. Kwasny (1975). Interactive consulting via natural language. *Commun. ACM*, **8**.

Smith, L. C. (1980). Implications of artificial intelligence for end user use of online systems. *Online Review*, **4**, 383.

Waltz, D. L. (1978). An English language question answering system for large data bases. *Commun. ACM*, **21**.

Waterman, D. A. (1978). Exemplary programming in RITA. In D. A. Waterman and F. Hayes-Roth (Eds), *Pattern Directed Inference Systems.* Academic Press, New York, pp. 261–279.

Webber, B. and A. Joshi (1982). Taking the initiative in natural language data base interactions: justifying why. In *COLING-82, Proceedings 9th International Conference on Computational Linguistics*, Prague, J. Horechy (Ed.). North-Holland Linguistic Series 47, Amsterdam, pp. 413–418.

Winograd, T. (1980). What does it mean to understand natural language? *Cognitive Sci.*, **4**, 209.