Cross Language Retrieval - English / Russian / French

A Working Paper for presentation at

AAAI 1997
American Association for Artificial Intelligence
Spring Symposium Series
March 24 - 26, 1997
Stanford University, California

System background information and research agendas

Marjorie M.K. Hlava
President, Access Innovations, Inc.
P O Box 8640
Albuquerque, NM 87198
mhlava@accessinn.com

Dr. Gerold Belonogov
Professor and Head of Department, VINITI, Moscow, Russia

Dr. Boris Kuznetsov
Head of Department, VINITI, Moscow, Russia

Richard Hainebach
Director, EPMS bv-Ellis Publications, The Netherlands
Introduction

In today's shrinking world, it is becoming evident that there is a large body of information and research available only in the language of the primary researcher, and much information cannot be shared between research communities without considerable translation and time delay. We would like to see these areas of research brought together to build a multilingual (not just bilingual) production, distribution, real-time translation, and retrieval system, with interfaces for each language of potential user communities. A prototype can be built using a few languages in differing character sets, and then additional languages can be added following the prototype system. In order to achieve this goal, we must build on existing research from around the world and create a worldwide research initiative to merge research lines. This will result in a "next generation" complete information system.

In small businesses (such as Access Innovations, Inc.) which compete with international, low-cost labor forces, we must continue to find ways to be more efficient and to ensure high-quality, very consistent results. At the same time, we find ourselves dealing daily with topics ranging from labor laws and employee benefits to water resources, chemistry, and medicine. To serve these twin masters of competition and varied subject areas, we have learned to depend heavily on natural language processing techniques. In the last six years, this approach has become even more important with the addition of valuable information resources in non-English languages and non-Latin alphabets.

This paper brings together discussions of three distinct lines of research, each of which has resulted in a working software product for the database production environment: machine aided indexing, a state-of-the-art translation system, and a multilingual search and retrieval system and interface.

1) The MAI - Machine Aided Indexing software developed by Access Innovations, Inc. produces proposed indexing terms from one or several knowledge bases. Each knowledge base is itself a database of text recognition rules. The knowledge base may be in any language or character set but it must match the target language. Current implementations index English, French, and Russian, and a Dutch rule base is under construction. The software has been adapted in English and French for an experiment with the multilingual documents of the European Parliament, and also in Russian and English for use with the Access Russia databases.

2) The machine translation systems RETRANS and ERTRANS were developed by the team of Dr. Gerold Belonogov at VINITI (the All Russian Institute for Scientific and Technical Information). RETRANS is a Russian-to-English translation system, and ERTRANS an English-to-Russian translation system. A French-to-Russian and Russian-to-French version of the translation software developed by the VINITI team also exists.
3) A multilingual search interface for Russian-to-English and English-to-Russian searching of target databases has been developed under Dr. Boris Kuznetsov, also of VINITI. This program has been named BROWSER. BROWSER searches databases in languages other than the input query language, building on the translation systems and adding software interrogation and relevance ranking of search results, for an interactive multilingual search front end.

Each of these three systems - the Access MAI, the RETRANS/ERTRANS translation software, and the BROWSER system - is based on a dictionary or rule base that creates a basic knowledge base for the system to weigh against text and present compatible word units in each of the language pairs for use by the reader. All systems allow editing of the output, and all systems will present the user with optional choices when they exist. All systems also allow weighting of the system output based on the subject matter of the input text ("plasma" in medicine vs. "plasma" in physics, for example), although they do it differently. Each of the systems is currently being used in several installations.

The three systems have also used multiple character sets (Latin and Cyrillic) without transliteration for the production of the final output in the source and target languages as well as in the source and target character sets. A description of the current software, platform features, etc. is attached as Appendix A.

The systems listed above are already paired with OCR systems to "Xerox into English," or to index full text directly from the source documents. Many other systems are connected as well.

We will suggest the next level of research and development for each product individually and for parallel research and development to bring together the three individual parts into a working, expandable, new software system to serve multiple character sets, multiple languages, and multiple subject sets worldwide.
SECTION I - THE THEORETICAL BASIS OF THE SYSTEMS

This section describes each of the three systems in general terms. Additional and in-depth information is available for each, and we would be pleased to provide demonstrations to interested parties.

I.A. THE MAI - An Overview

Machine Aided Indexing was developed to save time and enhance consistency for indexers processing multiple topics. It also extends the reach of indexers, increasing the kinds of items and the breadth of topics they can cover in an average work day. The Access MAI is based on a model first put into practice by the American Petroleum Institute (see references), a fairly simple and pragmatic algorithm using word matching, Boolean word phrases, proximity, adjacency, location, and other natural language processing techniques. Output selection for full text may invoke a relevance ranking system to limit the number of index terms selected, an especially important feature for full-text applications. The MAI system has three major components: 1) the Rule Builder, 2) the MAI Engine, and 3) the Statistics Package.

1) RULE BUILDER

A rule has three major components: the text string, conditions, and suggested term. Each is shown and defined here as a field in the rules database.

a. TEXT STRING, or keyword, is the term against which the MAI engine attempts to match text in the input file. The text string may be set for varying lengths; the default is four terms.

b. CONDITIONS, or logic, are instructions to the MAI engine qualifying, accepting, or rejecting assignment of an indexing term based on Boolean logic, relevance ranking, and other logic. Right- and left-hand truncation is used in the word standardization section of the Rule Builder.

c. SUGGESTED TERM, or index term, is the approved indexing term to be assigned if the logic is true.

There are five types of rules, divided into two categories: "simple" and "complex."

Simple rules use no conditions. They use either the identity rule, where the suggested term is the same as the matched text, or the synonym rule, where the matched text is synonymous with the suggested term. Simple rule examples are:
IDENTITY Rule

//TEXT: land productivity
USE land productivity

SYNONYM Rule

//TEXT: GNP
USE Gross National Product

Complex rules use one or more conditions. If a key word or phrase is matched, then the MAI may assign one, many, or no suggested terms, based on rule logic. There are three complex rule types: proximity, location, and format.

PROXIMITY Conditions

near- within up to 250 words from the matched phrase, in the whole document or limited to the same sentence. The default is three words.
with- in same sentence
mentions- in whole document, or normally in the title, abstract, or text fields

LOCATION Conditions (any field can be set by the rule builder)

SAMPLE LOCATION RULES

in title- if matched text is in title
in text- if matched text is in abstract or text
begin sentence- if matched text is located at beginning of sentence
end sentence- if matched text is located at end of sentence

FORMAT Conditions

all caps- if text is all caps
initial caps- if matched text begins with a capital letter

Complex rule examples follow.

//TEXT: science
IF (all caps)
USE research policy
USE community program
ENDIF
IF (near “Technology” AND with “Development”)
USE community program
USE development aid
ENDIF
IF (near "Technology" AND with "Environmental Protection")
USE community program
ENDIF
IF (near "Technology" AND with "Regional Innovation" AND with
  "Development")
USE community program
USE common regional policy
USE technology transfer
ENDIF
IF (near "Technology" AND with "Strategic Analysis")
USE community program
ENDIF

Machine Aided Indexing also offers several other customizing features:

**Truncation** - left and right for matches to words and phrases.

**User Definitions in rules** - for example: Search Language field and set IN RUSSIAN to TRUE or FALSE. Rule used after the match on the text may contain IF (IN RUSSIAN) where IN RUSSIAN is a user-defined concept not built into the rule language.

**Comments in rules** - These are not processed by the MAI Engine but are instructive to the user or rule maker as to why the rule is as it is.

**Adjustable input and output file formats**

**Real-time indexing using Microsoft Windows** - .DLL file for incorporation into existing A&I systems. Results are achieved within two seconds on suitable PCs.

If MAI is to be a successful tool, it is important to build a rules database that will produce relevant and consistent index terms. General rule-building starts with an existing thesaurus, (i.e., number of lead terms, number and quality of synonyms, the currency of the thesaurus, etc.) as well as a working knowledge of the types of source documents to be indexed. It is important to analyze the documents by the types of language and vocabulary used in the documents themselves and by the structure of each document (i.e., whether it is fielded, whether it contains an abstract, and/or whether it is full text). Once a serviceable rules database is established, it can be implemented using the MAI Engine.
2) MAI ENGINE

The MAI engine is essentially a set of matching algorithms which apply the rules built to the test input and produce a list of suggested index terms for the indexer.

3) STATISTICS PACKAGE

This feature is used to measure the performance of the MAI by comparing its performance to indexing by humans. In addition to performance measurement, the statistics are essential for tuning the knowledge base. Statistics bring missed terms and "noise" terms to light and point to where they appear. Identification of the most frequent MISS and NOISE occurrences allows us to concentrate on solving the problems that cause the most errors, thereby producing the greatest improvement. Information gathered by the statistics package is used to create new rules and to modify existing rules in the database.

a. HITS - when the MAI engine generates an indexing term identical to an index term which would have been assigned by a human indexer;

b. MISSES - when the MAI engine fails to generate an indexing term which would have been assigned by a human indexer; and

c. NOISE - when the MAI engine generates an indexing term which is genuinely incorrect, out of context, or illogical. (In the case of ÉPOQUE project, which is further referenced in the bibliography, this should not be confused with terms generated by the MAI but not selected by the human indexer.) Some cases will have both relevant (good terms but not listed in hit category) and irrelevant (bad indexing) noise.

The MAI increases the productivity of the general indexing process. It also provides for more consistent and deeper indexing. Tests from one project clearly show that without any human intervention the Machine Aided Indexing (MAI) did as well as the human indexer. Used in concert with human indexers as originally conceived, the system can provide faster, more consistent, more economical, and better quality indexing.
L.B. RETRANS AND ERTRANS - A Major Advance in Machine Translation

RETRANS and ERTRANS are essentially mirror image systems. They have the same theoretical basis and use the same processing algorithms. The difference is in the dictionary: one is built for an English target language, the other for a Russian target language. We will discuss RETRANS in some depth. The same list of attributes is true for ERTRANS, as well as for the French version of this translation software.

The RETRANS system was designed for automatic or interactive translation of polythematic texts from Russian into English. The system can process texts from a broad spectrum of application domains: economics, politics, military affairs, business, mechanical engineering, electrical engineering, power engineering, automatics and radio electronics, computer science, transportation, building and construction, aeronautics, cosmonautics, biology, medicine, physics, chemistry, mathematics, astronomy, ecology, agriculture, geophysics, geology, mining, metallurgy, and others. The same dictionary is in use at all times, but the user may select a subset which will weight the term usage to the vernacular of a specific field of expertise. The user may also add a personal dictionary to the system.

In contrast to other computer-assisted translation systems, the RETRANS system looks at fundamental units of meaning (phrases) rather than separate words. These word combinations, short sentences, and phraseological word-combinations make it possible to more precisely convey the meaning of translated texts. The system dictionary includes about 950,000 dictionary entries and covers 97-99% of the source polythematic texts. More than 80% of the dictionary consists of word combinations and phraseological combinations. The supplementary machine dictionaries contain more than 100,000 entries. The dictionary for ERTRANS is currently at 1,050,000 terms. Interactive translation screens can be created and adjusted for specific users.

Linguistic tools created and applied within the framework of computational linguistics can arbitrarily be divided into two components: declarative and procedural. Declarative tools include dictionaries of language and speech units, texts, and various grammatical tables. The procedural component includes the software tools that handle the declarative elements.

The RETRANS System includes the following basic procedural tools and assumptions:

1) The system's dictionaries contain primarily word and phraseological combinations. Only about 20% of the dictionaries are single word listings.
2) The translation routine for converting text from one language into another first translates the equivalents for the word combinations and the phraseological combinations. It then translates the remaining words.

3) In the process of text translation, procedures of morphological analysis and synthesis of Russian and English words using an analogy principle play an important role.

4) RETRANS performs automated morphological analysis and synthesis of Russian words, and is capable of processing texts of any subject field and with any word stock, including the alteration of vowels and consonants in suffixes and other morphemes.

5) The system uses automated normalization procedures of Russian words and word combinations, using procedures of morphological analysis and synthesis of words such as lemmatization, or breaking the words into their word roots.

6) The system performs morphological analysis and lemmatization of English words.

7) Automated procedures of text-based dictionary compilation and automated linguistic processing of machine dictionaries of Russian words and word combinations are part of the program. RETRANS automatically compiles Russian-English dictionaries of words by using parallel texts. Computer-assisted procedures are used for compiling the machine dictionaries using bilingual texts.

8) The system recognizes keywords and word combinations included in its thesaurus when it encounters them in texts. These procedures use the techniques of automatic morphological analysis and synthesis of words.

9) A complex of more than 30 procedures, named “linguistic operating system,” includes procedures for compiling text-based word-form dictionaries and for their linguistic processing, including inversions, sorting, setting theoretical operations with dictionaries, representing dictionaries in a form convenient for visual control, and so on.

The above list represents some of the work done by the authors of this article in the field of procedural linguistic tools. Some of these tools could be used for machine translation without essential changes from prior experimental and commercial linguistic processors, while others required considerable additional work. It was also necessary to elaborate new procedures. In particular, the system of Russian-English phraseological translation required the development of a procedure for extracting word combinations from Russian texts, a procedure for building searchpatterns of selected word combinations, a procedure for conducting searches in the Russian-English machine dictionary, and a procedure for selecting translated equivalents for fragments of the source Russian text from among numerous variants found in the machine.
dictionary. The new procedures also included those dealing with semantic-syntactic analysis of Russian texts and semantic-syntactic synthesis of English texts, as well as with the arrangement of translation results.

The system of Russian-English phraseological text translation operates sequentially. First, morphological analysis of the source text is carried out, and, using its results, nominal and verbal word combinations and phraseological units are identified on the basis of local semantic-syntactic analysis. Then all the words of text are normalized, and search patterns of word combinations and phraseological units are built into sequences of normalized word forms included in the search patterns.

This process is followed by searches in the Russian-English machine dictionary. Search patterns of alphabetically arranged Russian words and word-combinations serve as inputs in the dictionary. Search patterns of Russian words and word-combinations extracted from text are also arranged alphabetically. Searches in the dictionary are conducted using the “sliding starting point” method (batch-ordered search method). Translated equivalents of words and word combinations of the source text accompanied by the numbers of these words and by their combinations are produced as search results. Translated equivalents are arranged in order of increasing numerical values of the numbers of words and the combinations accompanying them.

The next stage of translation is selection, for each source text fragment, of the translated equivalent or equivalent series. If a number of equivalents are indicated in the dictionary, preference is given to the equivalents (or their series) which cover longer extracts of the source text. Alternative translation variants are excluded.

Intermediate translation results are arranged in the form of the structure shown in Table 1. This structure includes a centrally placed vertical column of ordinal numbers of the words of the source text, flanked on the left by words of the source Russian text, and on the right by English equivalents of Russian words and word combinations.

I.C. BROWSER - Multilingual Search Interface

BROWSER is a multilingual search interface which allows the user to input a search query in one language and search a database in an entirely different language. The Cyrillic BROWSER, for example, is a bilingual information retrieval system which is capable of processing English queries in original Russian language databases (Cyrillic texts). BROWSER requires no special
search language; the system communicates in limited natural English, processing queries prepared using natural English by translating the query into Russian, searching Russian language databases, and translating the retrieved records from Russian into English.

BROWSER automatically generates a set of Boolean subqueries using terms (words or word combinations) extracted from the initial user query. An individual set of records is produced for each subquery. All sets are arranged in order of decreasing relevance, so the first ranks will contain the most relevant records. Search results are automatically translated into English for English-speaking users.

Where conventional ways of processing queries in the interactive mode cause some problems for end users, in BROWSER the natural language queries from the user are processed automatically into the command language of the target system.

A brief history of BROWSER's development may be instructive.

During work on the project, different kinds of information retrieval system architectures intended for search in large Russian-language databases were considered. The lack of multilingual information retrieval systems (IRS) supporting multiple character sets such as Cyrillic and Latin, plus the selection of available machine-readable databases in more than one language, made it imperative that a way be found to search data in a different language from that of the researcher. Several options were evaluated:

1) Translation of the Russian language database by professional interpreters before loading in the traditional online IRS system. This has the drawback of the expense of intellectual translation of a large volume of information.

2) Translation of the Russian language database with the help of an automatic Machine Translation (MT) system before loading into the traditional online IRS system. The perceived drawback here is the poor quality of translation; in many cases the end user needs to see records in the original language to get more precise translation (with the help of a professional interpreter). Saving original language records in the database (to overcome this defect) almost doubles the volume of the database stored.

3) Extraction of keywords from original language (Russian) records, translation of them (perhaps with the help of automatic MT facilities), and formation for each record of an additional (English) language field with additional (English) language keywords. After such processing, the database could be loaded in the traditional online IRS system. Only the additional (English) language keywords field would be used for searching. Other fields would be used for output. Although this option preserves the original (Russian) language records and allows searching with a relatively small increase in the size of the database, the user has very little new or additional
(English) language information about the records (only a set of keywords). This is especially uncomfortable when dealing with large full-text records.

4) Loading of original Russian language databases into the existing BROWSER system designed by the VINITI team. This was the option selected.

BROWSER components and configuration

BROWSER is a complicated system, containing a large number of programs, files, directories, databases and other components, organized in three main sections:

1) Automatic translation of queries from English into Russian (ERTRANS),

2) Retrieval from Russian-language databases using Russian queries,

3) Translation of the retrieved results from Russian into English (RETRANS).

The main BROWSER directory and its subdirectories contain system programs and files. BROWSER works with four main types of files: queries, results, databases, scripts. The subdirectories Queries and Results store the input and output files; Database stores BROWSER databases; and Scripts holds the scenario files.

Query processing procedure

Queries are processed through the BROWSER programs and files using the following sequential procedures.

1) Analyze the natural English language query and extract search terms (words and word combinations).

2) Form the query as a set of terms.

3) Translate the query automatically from English into Russian with the help of the ERTRANS translation system.

4) Create the initial search statement using translated Russian terms.

5) Process the search statement in the specified database.

6) Generate the next search statements.
7) Estimate search results.

8) If the result is satisfactory, make the final output. If not, generate the next search statement.

9) Create the output results according to the script of the query processing.

10) Translate results automatically from Russian into English with the help of RETRANS translation system.

The BROWSER system provides a powerful, easy-to-use retrieval method to access information in Russian language databases. The system has many advantages:

1) There is no need to translate the database into English before loading it into the IRS.

2) The end user need not know Russian to conduct a search.

3) The end user need not know the command language of the IRS.

4) The quality of machine translation is high enough to assess relevance of retrieval.

5) All the stages of query processing are accomplished automatically, without the participation of an operator.

The value of the BROWSER search interface, along with the MAI Machine Aided Indexing and RETRANS and ERTRANS machine translation software, is clear. But further research and development is indicated to optimize the system’s usefulness.

SECTION II - INDIVIDUAL RESEARCH AGENDAS

II.A. Further Development of the MAI

To enhance the MAI program, Access Innovations plans the following research.

1) Develop a system to automatically generate rules from the changes made by the indexers or editors when reviewing the MAI indexing.

2) Apply the knowledge bases to the end user's query statement to produce an appropriate set of index terms to use in searching. This is an area for joint research with our Russian colleagues on the BROWSER team, so that search terms in one language can produce index terms in another.

3) Create a Web-based production system for remote locations using SGML and HTML coding and based on Internet protocols. This will create a truly worldwide virtual office environment.
II.B. Further Development of the RETRANS and ERTRANS Systems

The RETRANS and ERTRANS Systems form the conceptual basis for the development of many additional language translation systems. In order to speed the process for additional language systems without being fully tied to single pairings as is the traditional methodology, we suggest the following research agenda.

1) Generalize the procedures and dictionary structure in RETRANS, etc., i.e., separate the language-specific items from the non-language-specific.

2) Develop language-neutral conceptual schemas, where possible, so as to replace language-to-language processing with language-to-concept-to-language processing, allowing for one 11-language system rather than fifty-five language pair systems. This will be especially useful for European technical vocabularies.

3) Improve the procedures for semantic-syntactical analysis and synthesis of Russian and English texts in the RETRANS and ERTRANS systems.

4) Adjust general systems for high-quality translation of polythematic texts.

II.C. Research Agenda for the BROWSER System

We have identified a number of goals for further development of the BROWSER software.

1) Preserve good retrieval response time (seconds, dozens of seconds) in spite of drastic database volume increases. The response time for queries including multiple word combinations for short and long records (up to 1 MB) should be in the same range.

2) Provide three types of output:
   a. Full records,
   b. Relevant paragraphs (the paragraphs of records which contain the terms of the query),
   c. Relevant sentences (the sentences of records which contain the terms of the query).

3) Provide:
   a. Ranking of full records according to the level of relevance.
   b. Ranking of abridged records (including only relevant paragraphs) according to the level of relevance.
   c. Ranking of relevant paragraphs (not records) of all records according to level of relevance (hypertext output).

4) Provide a highlighting option for all types of output (highlighting keywords of the query in output
5) Provide a translation option for all types of output.

6) Enable fast search in full-text records as well as in structured records (for example bibliographic records) or mixed records.

7) Provide multi-base search facilities. The search strategy and ranking procedure would be chosen by processing the query against the most relevant database of the BROWSER system, and would be used again for query processing against less relevant databases.

8) Design a pilot version of a system that would automatically address queries to relevant databases. The system would provide an automatic choice of the set of relevant databases for query processing, according to natural language query contents, in a multi-base environment.

9) Create a pilot system for searching names presented in transliterated form. The system would have to take into account different possible versions of transliteration for an original Cyrillic notation of names.

10) Successful research and development of these features will create a multilingual retrieval system. The system would initially translate results into English only, and all language queries would initially be presented as concepts; natural language sentence queries would be a later step in the process. Of course, powerful concept dictionaries are needed for translation of concepts in each of the languages covered by the system, and we want to find and adapt as many of these as possible.

**SECTION III - THE COMPLETE SYSTEM RESEARCH AGENDA**

Recent research agendas have included exploring the expansion from language pairs to up to eleven output languages from a single input stream, presented in mixed character sets and expanded ASCII, using UNICODE, CCCII, and other algorithms. This will require adapting or creating a significant number of dictionary-based collections and moving them into knowledge bases. To create solid indexing and translations, these bases will need to include semantic, morphological, syntactical, and phraseological system applications, with relevance-ranked output evaluations from occurrence and mapping results.

Other areas will benefit from ongoing research efforts as well.

1) Individual improvements can be made to each of the software systems and their maintenance: a dictionary or rule base must change as the vernacular changes.
2) To bring together these systems to create a seamless multilingual database system, we must identify and learn to adapt or create rule bases and dictionaries for as many source and target languages as are needed by the user community.

3) Developing interfaces to existing database systems that transmit translated search queries to the database and translate the output back to the user is essential to creating a multilingual information retrieval system.

4) Related research initiatives to be pursued include the writing of calls for 1) OCR packages to seamlessly transfer data into the system, and 2) thesaurus management systems related to the translation and indexing systems to enhance concept translation between language systems.
Conclusion

We envision these three interlocking systems providing real-time interactions so that end users can query, in their own language, any document in any language and immediately view the results in their native tongue. That is, a Greek speaker who wants to read a machine-readable Finnish document would have only to enter a Greek language query into the system. The system would search for the document, translate it to Greek, and display the results in Greek. The result would currently be a rough translation which is “good enough” for the requester to get the gist of the article and glean the information necessary to make a decision and move forward, or could be refined by a translator for broader distribution and consideration. With future research efforts, the quality of translation will continue to improve.

If we are able to remove the language barrier for existing document collections, in all languages, in print or electronic form, cross-cultural communication will be greatly enhanced. International communication will result in more cooperation and collaboration, raising the level of global knowledge and facilitating implementation of research results to increase productivity and to further potentially beneficial scientific and technical discoveries.
APPENDIX A - SYSTEM REQUIREMENTS

- Machine Aided Indexing - MAI

The system runs on personal computers (IBM PC / AT286, 386, 486 and Pentium).
Operating System: MS-DOS
Rate of document processing: 56 pages per minute
Code written in: "C"
Working memory capacity: 580 KB minimum
Hard disk memory capacity: depends on file size - 5 KB minimum - Library of Congress Subject headings file is 200 MB; Science rule base is 15 MB
Type of input files: text files in ASCII
Size of input files: variable length - size dependent on machine memory

- RETRANS

The system runs on personal computers (IBM PC / AT286, 386, 486 and Pentium).
Operating System: MS-DOS 6.0 and higher
Rate of text translation in automatic mode on a 486: 500 standard typed pages (2000 characters) per minute (30-50 words / sec.).
Code written in: "C"
Working memory capacity: 580 KB
Hard disk memory capacity: 45 MB
Type of input files: text files in ASCII
Size of input files: not to exceed 150 KB at once

- ERTRANS

The system runs on personal computers (IBM PC / AT286, 386, 486 and Pentium).
Operating System: MS-DOS 6.0 and higher
Rate of text translation in automatic mode on a 486: 500 standard typed pages (2000 characters) per minute (30-50 words / sec.).
The code is written in "C"
Working memory capacity: 580 KB
Hard disk memory capacity: 47 MB
Type of input files: text files in ASCII
Size of input files: not to exceed 150 KB at once
- BROWSER

The system runs on personal computers (IBM PC / 386, 486 and Pentium).
Operating System: MS-DOS 4.0 and higher
Code written in: “C”
Working memory capacity: 590 KB
Hard disk memory capacity: 50 MB
Total free hard disk space for running the system: 5 MB for output files
Type of input files: text files in ASCII
Size of input files: not to exceed 150 KB at once
Size of output files: (size = number of queries* expected number of recall records* average size of record).

Total free disk space for running the system must not be less than 5 KB.
BIBLIOGRAPHY


