PSEUDO PARSING SWIFT-ANSWER ALGORITHM
by

(c) S Pal Asija 1989
Patent Attorney & Professional Engineer
7 Woonsocket Ave, Shelton, Conn. 06484
PH:(203)-736-9934 or 736-0774

INTRODUCTION

Pseudo parsing SWIFT-ANSWER algorithm is a subset of patented (4,270,182) SWIFT-ANWER algorithm which is based on the first pure software algorithm patent ever issued anywhere in the world. It is also a federal Trademark registered in the principal register of the Unites States Patent and Trademark Office. It is an acronym which stands for Special Word Index Full Text Alpha Numeric Storage With Easy Retrieval. It is called Pseudo parsing because it deviates substantially from conventional parsing algorithms, even though it accomplishes confusingly similar objectives. It is not a software package nor a key word search system.

NOT A KEYWORD SYSTEM

Some AI(Artificial Intelligence) experts and computational linguists have erroneously perceived this system as a keyword system and therefore have evaluated and crticized it as such. But in reality it is not a keyword system. In fact the system never asks the user for the keywords. Keywords if any are automatically created and managed by this system. It is strictly internal to the system and therefore completely transparent to the user.

Just as in human to human communications in this human/machine communication system also, neither the machine nor the human being is conscious of any keywords. When you ask a human being a question he or she does not ask you for key words, even though the respondant may subconsciously select and use some keywords to properly respond to you. Just as the user does not care what the subconscious of the respondant does, the user also does not care what the internal software of the system does to properly respond to the users communications.

BRIEF DESCRIPTION OF THE DRAWING

a) Figure 1 shows the program flow chart of the SWIFT-ANSWER algorithm.

b) Figure 2 shows the SWIFT-ANSWER Data Flo Diagram.

c) Figure 3 shows three performance curves as follows.

(i) Figre 3 (A) maps size of the polarized data base as a percentage of the original data base alng the Y axis as the structure of the data base is varied from fully structured to free form along the X axis for a constant data base size.

(ii) Figure 3 (B) maps size of the polarized data base along the Y axis as a function of the size of the original data base along the x axis for a data base of contstant structure.

(iii) Figure 3 (C) shows access time above a certain threshold is independent of the size of the original data base.

PSEUDO-PARSING ALGORITHM

The Pseudo-parsing SWIFT-ANSWER algorithm of this invention comprises the following steps.

a) Separation of a natural language text into senteneces, phrases and words.

b) Separation of Words into non-context and context words.

c) Separation of non-context words into noise words such as pronouns and prepositions and common words such as common words appearing too frequently in a file or data base.

d) Alphabetizing all context words.

e) Mapping frequency and location of all non-context words with respect to source data original files.

NOTE: The above mentioned five steps of the Pseudo parsing algorithm are applied to the natural language unstructred files in the batch mode and then again to the spontaneous convoluted questions in the real time mode.

f) performing mathematical operations such as taking highest common factor and lowest common multiple of the statistical information that correponds to context words in the question.

Note: This step in turn generates a lsit of prioritized answer which specified that the best answer based upon the totality of this question begins on disc so and so ,sector so and so and is so many bytes long and the second best answer begins on disc so and so, sector so and so and is so many bytes long and so on.

g) applying the user transparent boolean logic to different permutations and combinations of the context words in the question.

NOTE: The pseudo parsing is completed at this step. The remaining steps described in the patent deal with fetching and presenting the right answer in the right format to the user.

UNIQUE ALGORITHM

The algorithm is unique compared to the prior art because it is the only software that responds to a users erroneous spontaneous questions primarily because it performs the following user transparent functions automatically.

a) Automatic LIUs (Logical Information Units)
b) Automatic & Unlimited Dictionary.
c) Automatic Key Words
d) Automatic Boolean Logic
e) Automatic Prioritizing of Answers
f) Automatic Fault Tolerance
g) Automatic Context Determination.
h) Automatic DBM (Data Base Management)

The following functions and features are not automated.

a) Questions & Reframing of questions
b) Interpretation of Answers
c) Specification of the USER environment
d) Creation of the Unstructured Source Data Base
e) Selection of Special Features
f) Inputting of additional context dependant common words and 'synonyms & antonyms' ie searchonyms.

This unique algorithm creates the illusion of artificial intelligence without even using a conventional "Spell-Check" dictionary let alone spoon feeding rules of parsing, grammar, programming and knowledge of the world. The artificial intelligence if any in this system is inherent in the structure of the algorithm which makes it independent of the knowledge domain of the data base.

FIVE PAHSE ALGORITHM

a) Installation phase during which the computer asks you a series of questions to get to know your computer environment and your applicational needs including your data bases, so that it can load appropriate operating system, interface modules and drivers.

b) Batch Phase during which the algorithm pre-processes each file specified in phase (a) and extracts certain things from them.

c) Real Time Phase during which the same algorithm is applied to the question as was applied to the files in phase (b).

d) Priority phase during which the prioritizing algorithm ranks all possible answers without going to the data base files.

e) Presentation Phase during which the algorithm presents answers in the order established in the priority phase (a) supra by fetching them from the original files.

POWERFUL ALGORITHM

The power of the algorithm can be traced to the following precepts and principles.

a) All natural languages saturate. As a language saturates and the data base grows larger the probability of a new word appearing goes down and the proabability of a repeat word goes up. The length of the index does not increase although volume of cross-indexing does.

b) Some words such as most pronouns and propositions do not mean much even to people let alone computers
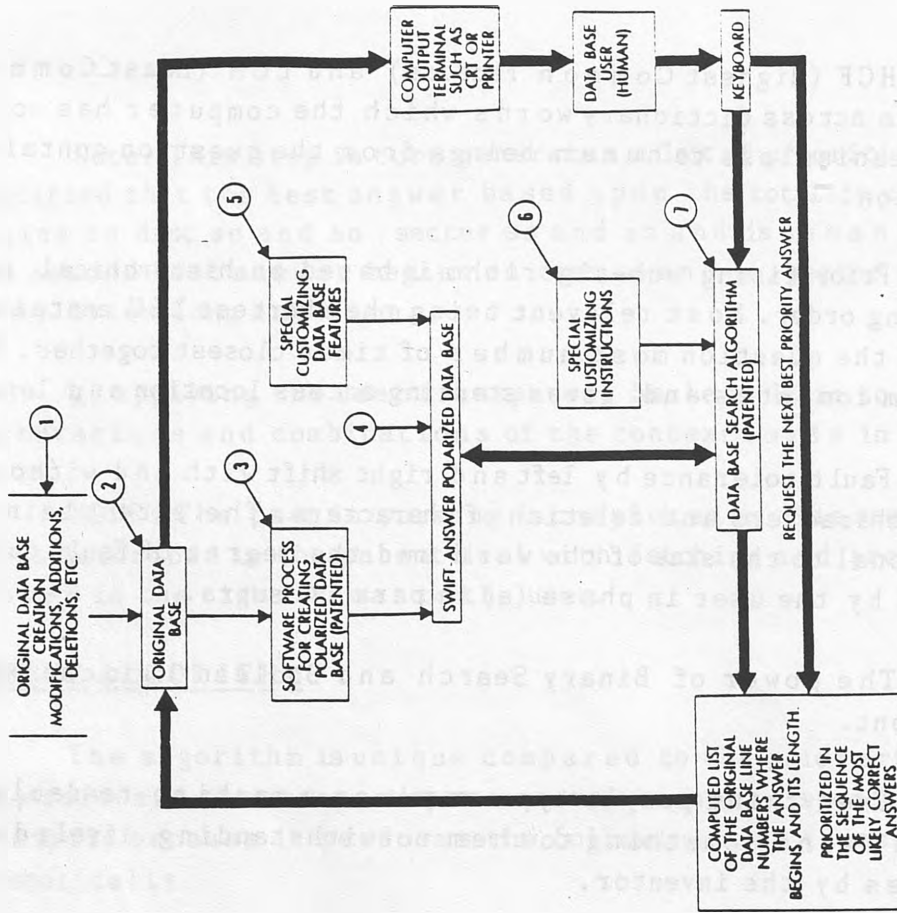
SWIFT-ANSWER DATA FLOW DIAGRAM

COMPUTER OUTPUT TERMINAL SUCH AS CRT OR PRINTER

DATA BASE USER (HUMAN)

KEYBOARD

ORIGINAL DATA BASE CREATION MODIFICATIONS, ADDITIONS, DELETIONS, ETC. ①

ORIGINAL DATA BASE ②

SOFTWARE PROCESS FOR CREATING POLARIZED DATA BASE (PATENTED) ③

SPECIAL CUSTOMIZING DATA BASE FEATURES ⑤

SPECIAL CUSTOMIZING INSTRUCTIONS ⑥

SWIFT-ANSWER POLARIZED DATA BASE ④

DATA BASE SEARCH ALGORITHM (PATENTED) ⑦

REQUEST THE NEXT BEST PRIORITY ANSWER

COMPUTED LIST OF THE ORIGINAL DATA BASE LINE NUMBERS WHERE THE ANSWER BEGINS AND ITS LENGTH.

PRIORITIZED IN THE SEQUENCE OF THE MOST LIKELY CORRECT ANSWERS

Figure 2

DATA FLOW DURING THE SWIFT-ANSWER INSTALLATION (ONE TIME)

DATA FLOW DURING THE SWIFT-ANSWER OPERATION

INPUT

COMMON WORDS 46

INPUT

SPECIAL FEATURES 50

COMPARE 70

ALPHABET-IZE IF NO MATCH 80

DISCARD COMMON WORDS 72

COMPARE UNIQUE WORDS WITH TEXT INDEX & COPY X-REF 82

PRESENT MOST CITED PARAGRAPH 30
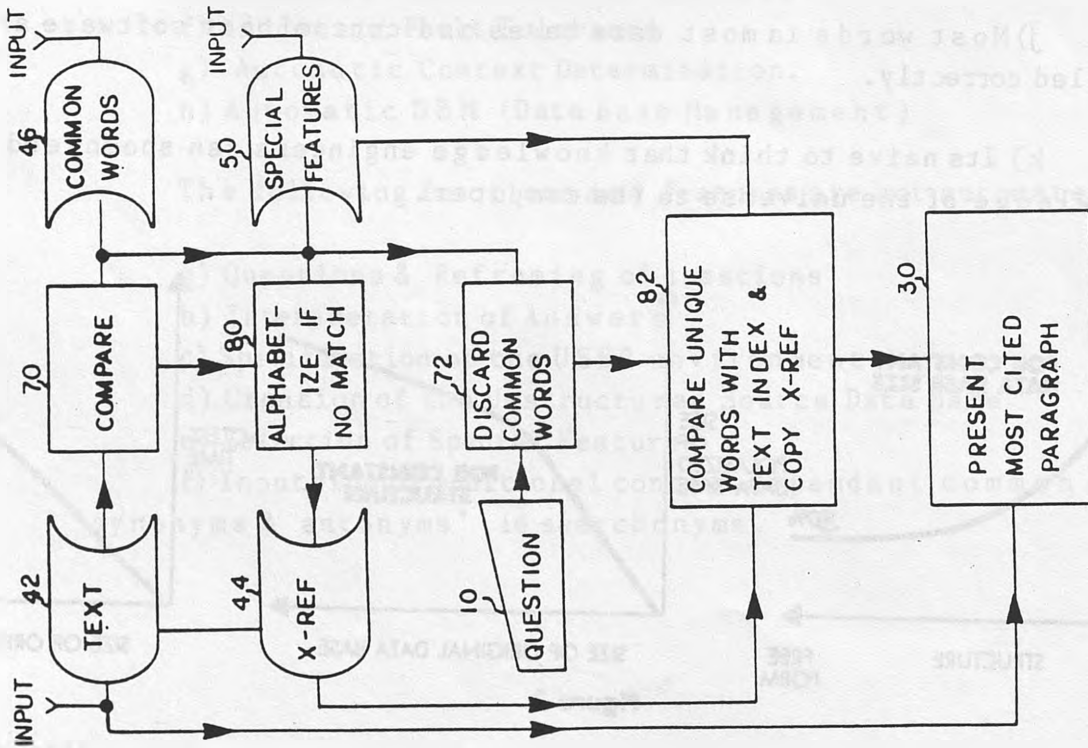
TEXT 42

X-REF 44

QUESTION 10

INPUT

Figure 1

c) HCF (Highest Common Factors) and LCM (Least Common Multiple) of all data across dictionary words which the computer has not been pre-told as meaningless to human beings from the question contain valuable information.

d) Prioritizing sub-algorithm is based on hierarchical relevence of decreasing order. Most relevent being the shortest LIU containing all words of the question most number of times closest together. The algorithm computes and gives starting access location and length.

e) Fault tolerance by left and right shift with and without addition of dummy characters and deletion of characters. The extent being proportional to the size of the word and the degree of fault tolerance specified by the user in phase (a) in para 14 supra.

f) The power of Binary Search and Boolean logic can be made user transparent.

g) Everything people type or put on a machine readable media probably means something to them notwithstanding lively demo given sometimes by the inventor.

h) Synonyms and antonyms both refer to the same context.

i) Von-Neumann serial computer is no match for the parallel processing brain which is not too well understood to begin with..

j) Most words in most data bases and concomitant software are spelled correctly.

k) Its naive to think that knowledge engineers can spoonfeed knowledge of the universe to the computer.
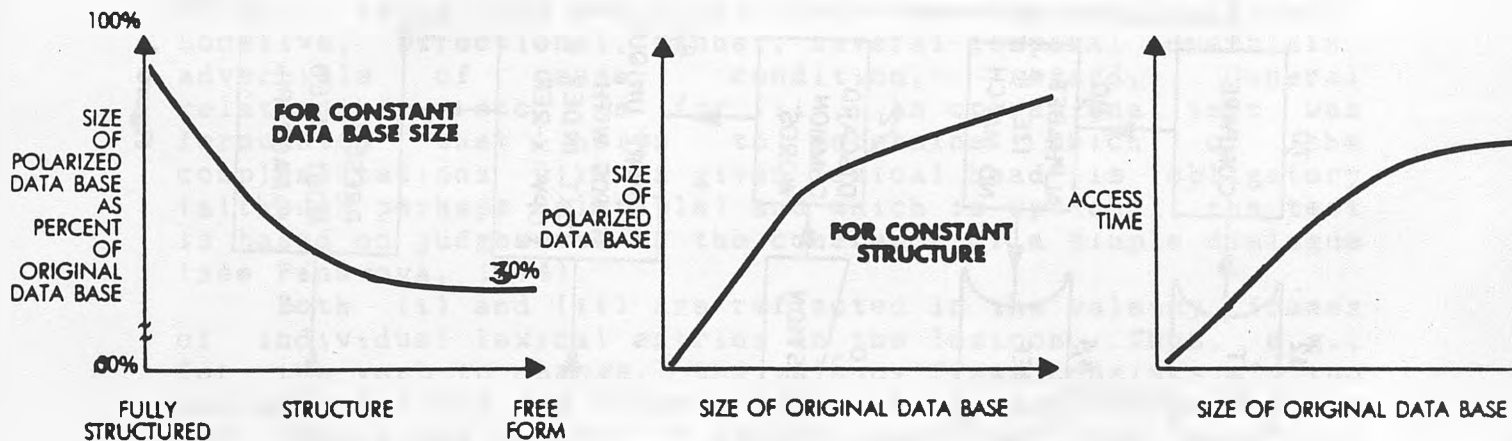

Figure 3