

1130 U.S. PTO
07/03/01

07-05-01

A

Attorney Docket No. YOR920010253US1

**IN THE UNITED STATES
PATENT AND TRADEMARK OFFICE**

PATENT APPLICATION

APPLICANT(S): Geoffrey G. Zweig
Mukund Padmanabhan

ASSIGNEE International Business
NAME: Machines Corporation

ASSIGNEE
RESIDENCE: Armonk, New York

TITLE: Information Extraction From Documents
With Regular Expression Matching

"Express Mail" Label No. EL793517350US

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated below and is addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231.

Date of Deposit: July 3, 2001

Signature: Denise A. Glaser

11046 U.S. PTO
09/898289
07/03/01

**ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231**

SIR:

Enclosed are the following papers relating to the above-named application for patent:

Application - 18 pages of Specification, 5 pages of Claims and 1 page of Abstract
2 Sheets of informal drawing(s)
Declaration and Power of Attorney (unexecuted)

CLAIMS AS FILED				
	NO. FILED	NO. EXTRA	RATE	CALCULATIONS
Total Claims	25-20 =	5	x \$18 =	\$90
Independent Claims	5-3 =	2	x \$80 =	\$160
Multiple Dependent Claim(s), if applicable			\$270 =	\$0
Basic Fee				\$710
			TOTAL FEE:	\$960

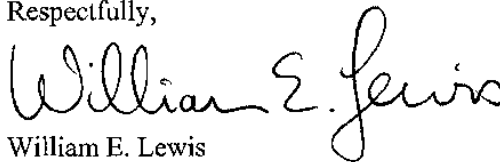
Please file the application and charge **International Business Machines Corporation Deposit Account No. 50-0510** the amount of \$960, to cover the filing fee. In the event of non-payment or improper payment of a required fee, the Commissioner is authorized to charge or to credit **International Business Machines Corporation Deposit Account No. 50-0510** as required to correct the error. Duplicate copies of this letter are enclosed.

Page 1 of 2

Attorney Docket No. YOR920010253US1

Please address all correspondence to: **Ryan, Mason & Lewis, LLP, 90 Forest Avenue, Locust Valley, New York 11560.** Telephone calls should be made to the under-signed attorney at (516) 759-2946.

Respectfully,



Date: July 3, 2001

William E. Lewis
Reg. No. 39,274
Attorney for Applicant(s)
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, New York 11560

Page 2 of 2

**INFORMATION EXTRACTION FROM DOCUMENTS WITH REGULAR
EXPRESSION MATCHING**

Field of the Invention

5 The present invention relates to data processing techniques and, more particularly,
to techniques for providing automatic information extraction from data streams such as
text and spoken documents.

Background of the Invention

10 There are many situations in which it is desirable to extract key pieces of
information from documents. For example, in transcribed voice mail messages, the name
of the caller and any return numbers that were left are crucial for summarizing the call.
Or when resumes are submitted to a company along with a cover letter, it is desirable to
extract the job objective and salary requirements of the applicant, in order to determine if
a suitable match exists. This invention comprises a simple, efficient, and effective way of
extracting such information.

15 In the past, several techniques have been developed to solve the somewhat
simpler problem of "named entity extraction." In this task, the goal is to identify all
occurrences of certain classes of words in a document. For example, all the
person-names, city-names, dates, and times might be identified. One way of identifying
such entities is to train a statistical classification system to tag each word in the document
20 as either a "person-name," "city-name," "date," "time," or "other" word. Examples of
this sort of approach are disclosed in, e.g., Ratnaparkhi, "A Maximum Entropy Part of
Speech Tagger," Conference on Empirical Methods in Natural Language Processing,
University of Pennsylvania, 1996; and Brill, "Transformation-Based Error-Driven
Learning and Natural Language Processing: A Case Study in Part of Speech Tagging,"
25 Computational Linguistics, December 1995, the disclosures of which are incorporated by
reference herein. While the problem of extracting key pieces of information can also be

viewed as a tagging problem, in which each word is tagged as either “key information” or “irrelevant,” such a tagging approach fails to explicitly identify the portions of text that are important.

Thus, there is a need for data processing techniques which explicitly identify portions of data that are sought to be identified, rather than only implicitly identifying, or tagging, such portions of data.

Summary of the Invention

The present invention provides techniques for exploiting the readily-identifiable structure of language to explicitly identify portions of data in a document that a user seeks to be identified, e.g., relevant or important information. It is to be understood that the term “document,” as used herein in accordance with the invention, is intended to broadly refer to a sequence of symbols (e.g., letters, numbers, characters, etc.), such as a data stream representative of text and/or spoken messages. However, it is to be understood that the invention applies to many input forms other than typed text or spoken messages. By way of further example only, the invention may also be applied to information extraction from data streams representing DNA (deoxyribonucleic acid) sequences, RNA (ribonucleic acid) sequences, amino-acid sequences, and audio and video sequences that are preprocessed into a discrete symbolic form.

More specifically, the present invention realizes that, in common linguistic usage, people typically convey information in highly stereotyped ways. So, for example, in a phone call, the caller is likely to identify himself in one of just a few ways, e.g., “Hi <recipient-name>, it’s <caller-name>” (e.g., “Hi John, it’s Bob”) or “<recipient-name>, <caller-name> here” (e.g., “John, Bob here”). Or in a cover letter, a job applicant is likely to express his interest with a phrase like “I am looking for a job in <field>” (e.g., “I am looking for a job in electrical engineering”).

In accordance with this realization, the present invention provides techniques for enumerating regularly identifiable or stereotypical phrases that people commonly use to

YOR920010253US1

convey particular information, and where exactly in these phrases the particular information is to be found. In one embodiment, such phrases are referred to as "regular expressions." Using such enumerated phrases, the invention is able to automatically identify them in an input data stream and then identify and extract the particular information associated with the phrase that is being sought, e.g., important or relevant information.

This approach has several advantages over the previously mentioned tagging approach. First, the phrases are readily expressible as regular expressions, and this allows them to be stored in one of several common programming languages, e.g., as a "flex" or "perl" program. Both these languages have built-in support for regular expressions. Secondly, the use of regular-expression technology enables extremely fast pattern matching and information extraction with highly optimized standard programs. Thirdly, the phrases can be identified without the expensive and time consuming step of gathering and annotating a large "training" database. Finally, if desired, the technique can be combined with statistical based procedures, e.g., by using the phrases as features in a statistical system, or by using a statistical procedure to automatically collect a large pool of candidate phrases which can then be selected for inclusion by a human.

One skilled in the art will realize various applications for the invention. By way of example only, such enumerated phrases may be automatically identified in a transcribed voice mail message, and the important information in the message, e.g., a phone number or caller's name, may be automatically extracted and presented to the user or stored for subsequent use.

These and other objects, features and advantages of the present invention will become apparent from the following detailed description of illustrative embodiments thereof, which is to be read in connection with the accompanying drawings.

Brief Description of the Drawings

FIG. 1 is a flow diagram illustrating an information extraction methodology according to an embodiment of the present invention; and

FIG. 2 is a block diagram illustrating a generalized hardware architecture of a data processing system suitable for implementing an information extraction methodology according to the present invention.

Detailed Description of Preferred Embodiments

As will be explained in detail below, the present invention provides techniques for using stereotypical phrases or “regular expressions” to identify the information-bearing portions of a document, and thereby explicitly identify the information found therein. Regular expressions, as have been discussed in Aho et al., “Compilers: Principles, Techniques, and Tools,” Addison-Wesley, 1985, the disclosure of which is incorporated by reference herein, comprise symbolic strings, typically word sequences, that can be constructed by the basic operations of “and” and “or” acting on an atomic alphabet of symbols. So, for example, the pattern “the dog” produces a match only when, in the data being considered, the symbol “the” is followed by the symbol “dog.” This represents the operation of concatenation or “and.” The pattern “the (dog | cat)” illustrates the “or” operation, i.e., the pattern matches either the string “the dog” or “the cat.” The operation of “classing” is also possible. So, for example, a “pet” class may be defined as either “dog” or “cat.”

`<animal> ::= (dog | cat)`

Then, subsequent regular expressions can make use of this class, for example, the pattern: “the <pet>.” A wide variety of extensions to regular expressions exist, for example, restrictions on where in a document a pattern is found, or restrictions on the length of repeated sequences in a regular expression. In accordance with an illustrative embodiment of the invention, the term “regular expression” is taken to mean any form of pattern that is matchable in the *flex*, *lex*, or *perl* programming languages, although the

invention is not limited thereto. The programming languages *flex* and *lex* are described in J.R. Levine et al., "Lex & Yacc," 1992, the disclosure of which is incorporated by reference herein. The programming language *perl* is described in L. Wall et al., "Programming Perl," 1996, the disclosure of which is incorporated by reference herein. Similarly, in accordance with such an illustrative embodiment, the term "matching" refers to the processes by which these programs match their input.

In accordance with an embodiment of the invention, when an information-bearing regular expression is matched, an action is taken to present the information in an appropriate way. There are two main types of regular expressions that may be used in this particular embodiment: trigger prefixes and trigger suffixes.

A trigger prefix is a characteristic phrase that typically precedes a piece of information, for example, in a voice mail message, "give me a call back at" is a trigger phrase that precedes a phone number.

A trigger suffix is a phrase that typically follows a key piece of information, even when a trigger prefix is not present. For example, the phrase "talk to you later bye" is often a post-facto signal that a phone number has been provided, when the words immediately preceding the phrase are a sequence of numbers.

In this embodiment of the invention, trigger prefixes and suffixes may be used either separately, or combined together into larger composite patterns. The operation of word classing is also inherent in this embodiment of the invention. For example, a set of personal names such as those found in a phone book may be identified as a "person-name" class. Subsequently, patterns of the form "Hi, it's <person-name>+" can be matched, and the words matching the "person-name" tokens can be displayed as the caller. It is to be understood by those skilled in the art that there are numerous ways of combining trigger prefixes and suffixes, word classes, and other pattern-matching operations, such as the use of multiple states, each with its own set of matchable patterns (as may be done in *flex*) in order to identify and extract information.

In addition to being used as a stand-alone method for information extraction, regular expressions can be in combination with statistical techniques. For example, the existence of a pattern such as "Hi <person-name>, it's <person-name>" can be used as a feature in a maximum-entropy word-tagging system.

Referring now to FIG. 1, a flow diagram illustrates an information extraction methodology according to an embodiment of the present invention. As shown in FIG. 1, the methodology 10 begins by obtaining the input data 12 to be processed. As mentioned above, the input data is a sequence of discrete symbols or a document, such as may be representative of text and/or a transcribed spoken message.

In step 14, the input data is normalized. This process may include standardizing any characteristics of the input that are irrelevant to, or a hindrance to, information extraction. An example of normalization, for text input, is to change all the letters in the text stream to lower-case and to replace multiple blank spaces between words by a single blank space. This makes subsequent pattern matching easier.

By way of example, the following is a sample *flex* program for normalizing text input:

```
%{  
  
#include <iostream.h>  
#include <ctype.h>  
  
%}  
  
%option noyywrap  
%option never-interactive  
  
ws [\t]+  
  
%%  
  
- {cout << " ";}  
[!\\.,\"()?] {}
```

YOR920010253US1

6

```

\<[^\>]*\>{ws}*      {}
{ws}                  {cout<<" ";}
.                      {cout<<char(tolower(*yytext));}

```

```
%%
```

```

5  main()
   {
       yylex();
   }

```

In step 16, one or more previously-stored word (or, in general, one or more symbols) lists are used to identify and mark occurrences of words that belong to certain classes. By way of example, such a class-annotation operation may include marking all proper names by prefixing them with an “!” mark.

By way of example, the following is a sample C++ program for annotating names identified in a normalized text stream:

```

15  #include <iostream.h>
    #include <fstream.h>
    #include <stdlib.h>
    #include <set>
    #include <string>

20  main()
    {
        set<string>name;
        ifstream namesin("../data/person_names"); // a text file listing names
        if (!namesin)
25      {
            cout << "Unable to find ../data/person_names\n";
            exit(1);
        }

        while (!namesin.eof())
30      {
            string s;

```

```

        namesin >> s >> ws;
        name.insert(s);
    }
    namesin.close();
5    while (!cin.eof())
    {
        string s;
        cin >> s >> ws;
        if (name.count(s))
10         cout << "!" << s << " ";
        else
            cout << s << " ";
    }
}

```

15 In step 18, the normalized and class-annotated text is read by a regular-expression matching program, and the information-bearing portions of the text are analyzed and extracted. In a preferred embodiment, the regular expressions are specified in and matched by a specialized regular expression matching program, such as may be implemented in *flex*. In this process, it is understood that all the mechanisms available in

20 such a programming language may be used. These include the use of special start states, backtracking, context-sensitive matching, and embedded special-purpose "C" code.

Specifically, the matching operation in step 18 takes the normalized, class-annotated input data and identifies the regular expressions in the data by comparing the data to previously-stored regular expressions. These previously-stored regular

25 expressions may include, for example, trigger prefixes, trigger suffixes, word classes, and other matchable patterns. When a match occurs, the structure of the regular expression is analyzed to identify the information-bearing portion of the expression. The information is then extracted from the expression.

By way of example, the following is a sample *flex* program for identifying caller

30 names and phone numbers in a transcribed voice mail messages. As is evident, the program enumerates a plurality of regular expressions, categorized as

PHONE_NUMBER_STATE, NAME_STATE, EXTENDED_NAME_STATE, and
TERMINATION_STATE, that are considered during the matching step 18:

```
%{  
  
5  #include <iostream.h>  
   #include <string.h>  
  
   const char *number_type;  
   const char *tl = "tieline";  
   const char *extension = "extension";  
10  const char *pager = "pager";  
   const char *fax = "fax number";  
   const char *none = "";  
  
   char number_buff[10000];  
  
   %{  
  
15  %option noyywrap  
   %option never-interactive  
  
   /* many patterns like IN-GENERIC_NAME-AT-DIGIT+unaccounted for */  
  
   spelled_digit  
   ((zero | oh | one | two | three | four | five | six | seven | eight | nine | hundred | thousand | ten |  
   eleven | twelve | thirteen | fourteen | fifteen | sixteen | seventeen | eighteen | nineteen | twenty |  
20  thirty | forty | fifty | sixty | seventy | eighty | ninety | hundred)[ ]?)  
   numeric_digit ((1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10)[ ]?)  
   digit ([ ]?{spelled_digit} | [ ]?{numeric_digit})  
   phone_number ((country[ ]code[ ]{digit}{2,})?(toll[ ]free)?(area[ ] | area[ ]code[ ]  
   )?{digit}{2,} ((or[ ] | that's [ ] | sorry[ ] ) | {digit}{2,})?(option[ ]{digit}{+})?(extension[  
25  ]{digit}{2,})? tie_line ([ ]?(tie[ ]line | tieline)([ ]number)?  
   name ![a-z'] +  
  
   %x PHONE_NUMBER_STATE  
   %x NAME_STATE  
   %x EXTENDED_NAME_STATE  
30  %x TERMINATION_STATE
```

YOR920010253US1

9

"give me buzz at" |
 "give me a holler" |
 "give me a holler at" |
 "give me a ring" |
 5 "give me a call" |
 "give me a ring at" |
 "give me a call at" |
 "give me a call at home" |
 "give me a call at work" |
 10 "give me a call at your convenience" |
 "give me a call back at" |
 "give me a call back on" |
 "give me a call back with that information at" |
 "give me a call if you get a chance" |
 15 "give me a call please" |
 "give me a call please at" |
 "give me a call when you get a chance" |
 "give me a call when you can" |
 "give me a call when you can please" |
 20 "give us a call" |
 "give us a call at" |
 "here" |
 "i can be reached at" |
 "i am at" |
 25 "i'm at" |
 "i'll be at" |
 "i'm on" |
 "leave me a message" |
 "leave me a message at" |
 30 "my number is" |
 "my number here is" |
 "my number here in "(?![a-z]+[])"is" |
 "my number's" |
 "my phone number's" |
 35 "number again is" |
 "number at home is" |
 "number at work is" |
 "number here is" |
 "number is" |
 40 "number's" |

YOR920010253US1

11

```

"outside line" |
"outside" |
"please call" |
"phone" |
5 "please return our call at" |
"reach met at" |
"returning your call" |
"talk to you later" |
"try me at" |
10 "try me at this number" |
"you could reach us at" |
"when you have a chance" |
"when you get a chance" {number_type=none; BEGIN PHONE_NUMBER_STATE;}

"fax number is" |
15 "my fax number's {number_type=fax; BEGIN PHONE_NUMBER_STATE;}
"my pager" |
"my pager is" |
"give me a page" |
"give me a page at" |
20 "page me at" |
"page me" |
"page number" {number_type=pager; BEGIN
PHONE_NUMBER_STATE;}

<PHONE_NUMBER_STATE>{phone_number} {cout << "Number: " << number_type
25 << yytext << endl; BEGIN INITIAL;}
<PHONE_NUMBER_STATE>. {unput(*yytext); BEGIN INITIAL;}

"it's "{phone_number} |
"on "{phone_number} {char *c=strchr(yytext, ' '); c++; cout << "Number: " << c <<
endl;}

30 "it's me" |
"it's just me" {cout << "Name: me" << endl;}
"this is your wife" {cout << "Name: wife" << endl;}
"this is receiving" {cout << "Name: receiving" << endl;}
"this is the mail room" {cout << "Name: the mail room" << endl;}

35 ^({name}[ ]){2,3} {cout << "Name: " << (strchr(yytext, ' ')+1) << endl;}
("hi"|"hey"|"hello"){name}" "{name}" "({name})"? {char *c=strchr(yytext, ' ');

```

YOR920010253US1

12

```

c=strchr(c+1, ' '); cout << "Name: " << (c+1) << endl;}

"good morning " {name} " " |
"yeah" {name} " " |
({name}[ ])? "this is" |
5 {name} "it's" |
{name} "hi" |
"hi i'm" |
"my name is" |
"my name's" |
10 "the name is"
"the name's" |
"hey"({name}[ ])? "this"("is")? |
"hi"({name}[ ])? "this"("is")? |
({name}[ ])? "hi it's" |
15 ({name}[ ])? "hey it's" |
({name}[ ])? "how you doing it's" |
{name} "this" {BEGIN NAME_STATE;}

<NAME_STATE>({name}[ ])+ {cout << "Name: " << yytext << endl;
number_type=none;
20 BEGIN PHONE_NUMBER_STATE;}

<NAME_STATE>("me" | "mom" | "your sis" | "your sister" | "your wife" | "your brother" |
"your father" | "your husband") {cout << "Name: " << yytext << endl; BEGIN
INITIAL;}

<NAME_STATE>({name}[ ])+ "from" |
25 <NAME_STATE>({name}[ ])+ "calling from" |
<NAME_STATE>({name}[ ])+ "i'm calling from" |
<NAME_STATE>({name}[ ])+ "with" |
<NAME_STATE>({name}[ ])+ "i'm with" |
<NAME_STATE>({name}[ ])+ "at" |
30 <NAME_STATE>({name}[ ])+ "over at" |
<NAME_STATE>({name}[ ])+ "down at" |
<NAME_STATE>({name}[ ])+ "over in" |
<NAME_STATE>({name}[ ])+ "up in" |
<NAME_STATE>({name}[ ])+ "down in" |
35 <NAME_STATE>({name}[ ])+ "up on" |
<NAME_STATE>({name}[ ])+ "down on" |

```

```

<NAME_STATE>({name}[ ])+in" {cout << "Name: " << yytext; BEGIN
EXTENDED_NAME_STATE;}
<NAME_STATE>. {unput(*yytext); BEGIN INITIAL;}

```

```

5  <EXTENDED_NAME_STATE>"i'm" |
   <EXTENDED_NAME_STATE>"i" |
   <EXTENDED_NAME_STATE>"i've" |
   <EXTENDED_NAME_STATE>"i'd" |
   <EXTENDED_NAME_STATE>"my" |
   <EXTENDED_NAME_STATE>"it's" |
10  <EXTENDED_NAME_STATE>"it is" |
   <EXTENDED_NAME_STATE>"we" |
   <EXTENDED_NAME_STATE>"when" |
   <EXTENDED_NAME_STATE>"you" |
   <EXTENDED_NAME_STATE>"calling" |
15  <EXTENDED_NAME_STATE>"now" |
   <EXTENDED_NAME_STATE>"a" |
   <EXTENDED_NAME_STATE>"if" |
   <EXTENDED_NAME_STATE>"give" |
   <EXTENDED_NAME_STATE>"their" |
20  <EXTENDED_NAME_STATE>"and" |
   <EXTENDED_NAME_STATE>"notifying" |
   <EXTENDED_NAME_STATE>"returning" |
   <EXTENDED_NAME_STATE>"sorry" |
   <EXTENDED_NAME_STATE>"can" |
25  <EXTENDED_NAME_STATE>"got" |
   <EXTENDED_NAME_STATE>"earlier" |
   <EXTENDED_NAME_STATE>"your" |
   <EXTENDED_NAME_STATE>"just" |
   <EXTENDED_NAME_STATE>"that" |
30  <EXTENDED_NAME_STATE>"could" |
   <EXTENDED_NAME_STATE>"would" |
   <EXTENDED_NAME_STATE>"before" |
   <EXTENDED_NAME_STATE>"thanks" |
   <EXTENDED_NAME_STATE>"thank" |
35  <EXTENDED_NAME_STATE>"wanted" |
   <EXTENDED_NAME_STATE>"following" |
   <EXTENDED_NAME_STATE>"regarding" |
   <EXTENDED_NAME_STATE>"may" |
   <EXTENDED_NAME_STATE>"let" |
40  <EXTENDED_NAME_STATE>"he" |

```

YOR920010253US1

14

```

    <EXTENDED_NAME_STATE>"about" |
    <EXTENDED_NAME_STATE>"again" |
    <EXTENDED_NAME_STATE>"wanting" |
    <EXTENDED_NAME_STATE>"appreciate" |
5    <EXTENDED_NAME_STATE>"was" |
    <EXTENDED_NAME_STATE>"we're" |
    <EXTENDED_NAME_STATE>"or" |
    <EXTENDED_NAME_STATE>"you've" |
    <EXTENDED_NAME_STATE>"right" |
10   <EXTENDED_NAME_STATE>"first" |
    <EXTENDED_NAME_STATE>[ ]?"tieline" |
    <EXTENDED_NAME_STATE>[ ]?"extension" |
    <EXTENDED_NAME_STATE>" " {digit} |
    <EXTENDED_NAME_STATE>"please" {cout << endl; char *q=yytext+strlen(yytext);
15   while (q>yytext) unput(*--q); number_type=none; BEGIN
    PHONE_NUMBER_STATE;}
    <EXTENDED_NAME_STATE>. {ECHO;}

    .{;}

    %%

20   main()
    {
        yylex();
        cout << endl;
    }

```

25 Thus, by way of example, if the voice message input to the methodology 10 contains the phrase "call me back at 555-1234," the above program determines that the phrase "call me back at" matches one of the enumerated expressions. Then, the information associated with the phrase "call me back at," namely, "555-1234," is identified as the information-bearing portion of the expression, and extracted.

30 The extracted output data 20 may then be presented to a user. For example, this may include displaying the information to the user on a display or converting the information to audio and audibly presenting the information to the user in accordance

with an output speaker. The extracted data may also be used to take some other specified action, e.g., place extracted data in an electronic phone book, notify user of a message from a particularly important person, notify a user of a message on a particular topic, etc.

Referring now to FIG. 2, a block diagram illustrates a generalized hardware architecture of a data processing system suitable for implementing an information extraction methodology according to the present invention. As shown, the system 30 includes a data capture device 32, a data signal processor 34, a data output device 36, and memory 38.

The data capture device 32 generally obtains the data stream that is to be processed by the system 30. The device's specific structure and operations depend on the application in which the system is employed. For example, in a voice mail application, the data capture device may be a speech processing device capable of taking an audio stream recorded in accordance with a voice mail system and converting or transcribing the audio stream to a discrete data or symbol sequence. The device may, for example, utilize conventional speech recognition techniques or some other known transcription techniques. In a text based application, the device may be a digital scanner capable of converting a text document into a discrete data or symbol sequence. The device may, for example, utilize conventional optical character recognition techniques. Of course, depending on the data type, the data capture device may take on alternate forms.

The data signal processor 34 generally performs and/or controls the operations of the methodologies of the present invention, in conjunction with the memory 38. That is, the signal processor performs and/or controls the steps depicted in the flow diagram of FIG. 1.

It is to be appreciated that the term "processor" as used herein is intended to include any processing device, such as, for example, one that includes a CPU (central processing unit) or other processing circuitry. For example, the processor may be a digital signal processor, as is known in the art. Also the term "processor" may refer to one or more individual processors. The term "memory" as used herein is intended to

include any memory devices associated with a processor or CPU, such as, for example, RAM, ROM, a fixed memory device (e.g., hard drive), a removable memory device (e.g., diskette), flash memory, etc. It is to be understood that the previously-generated word classes and regular expressions used in the class-annotation and expression matching operations, described above in accordance with the invention, may be stored in memory 38 and accessed when required.

Accordingly, one or more computer software programs including instructions or code for performing the methodologies of the invention, as described herein, may be stored in one or more of the associated memory devices (e.g., ROM, fixed or removable memory) and, when ready to be utilized, loaded in part or in whole (e.g., into RAM) and executed by a CPU.

The data output device 36 generally presents the data extracted from the input data stream to a user. Again, the device's specific structure and operations depend on the application in which the system is employed. For example, in a voice mail application, the data capture device may be an audio output system capable of taking the extracted discrete data and converting the data to a form to be audibly output to the user. The device may, for example, utilize conventional text-to-speech or other speech synthesis techniques. Such techniques may also be used in a text based application. Alternatively, the device may be a display for visually presenting the extracted data to the user. Of course, depending on the data type, the data output device may take on alternate forms.

It is to be understood that the elements illustrated in the figures may be implemented in various forms of hardware, software, or combinations thereof, e.g., one or more digital signal processors with associated memory, application specific integrated circuit(s), functional circuitry, one or more appropriately programmed general purpose digital computers with associated memory, etc. For example, the information extraction methodologies of the invention may be implemented in various personal computing devices. Given the teachings of the invention provided herein, one of ordinary skill in the

related art will be able to contemplate other implementations of the elements of the invention.

5 Although illustrative embodiments of the present invention have been described herein with reference to the accompanying drawings, it is to be understood that the invention is not limited to those precise embodiments, and that various other changes and modifications may be effected therein by one skilled in the art without departing from the scope or spirit of the invention.

Claims

What is claimed is:

1. A method of automatically processing an input sequence of data symbols, the method comprising the steps of:

5 identifying at least one regularly identifiable expression in the input sequence of data symbols;

identifying at least a portion of information associated with the at least one regularly identifiable expression; and

extracting the portion of information.

10 2. The method of claim 1, wherein the regularly identifiable expression identifying step comprises comparing the input sequence of data symbols to one or more previously-stored regularly identifiable expressions to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions.

15 3. The method of claim 1, further comprising the step of normalizing the input sequence of data symbols prior to identifying the regularly identifiable expression.

4. The method of claim 1, further comprising the step of identifying one or more classes of data symbols in the input sequence prior to identifying the regularly identifiable expression.

20 5. The method of claim 1, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically precedes a particular portion of information.

6. The method of claim 1, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically follows a particular portion of information.

5 7. The method of claim 1, wherein the extracted portion of information is used to take a specified action.

8. The method of claim 1, wherein the extracted portion of information is at least one of visually and audibly presented to the user.

10 9. The method of claim 1, wherein the regularly identifiable expression identifying step is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language.

10. The method of claim 1, wherein the input sequence of data symbols is representative of at least one of text data, transcribed spoken data, deoxyribonucleic acid sequence data, ribonucleic acid sequence data, amino-acid sequence data, audio sequence data, and video sequence data.

15 11. The method of claim 1, wherein the input sequence of data symbols is representative of a voice mail message.

12. Apparatus for automatically processing an input sequence of data symbols, the apparatus comprising:

20 at least one processor operative to: (i) identify at least one regularly identifiable expression in the input sequence of data symbols; (ii) identify at least a portion of information associated with the at least one regularly identifiable expression; and (iii) extract the portion of information; and

YOR920010253US1

20

memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations.

5 13. The apparatus of claim 12, wherein the regularly identifiable expression identifying operation comprises comparing the input sequence of data symbols to one or more regularly identifiable expressions, previously stored in the memory, to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions.

10 14. The apparatus of claim 12, wherein the at least one processor is further operative to normalize the input sequence of data symbols prior to identifying the regularly identifiable expression.

15 15. The apparatus of claim 12, wherein the at least one processor is further operative to identify one or more classes of data symbols in the input sequence prior to identifying the regularly identifiable expression.

16 16. The apparatus of claim 12, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically precedes a particular portion of information.

17. The apparatus of claim 12, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically follows a particular portion of information.

20 18. The apparatus of claim 12, wherein the extracted portion of information is used to take a specified action.

19. The apparatus of claim 12, wherein the extracted portion of information is at least one of visually and audibly presented to the user.

20. The apparatus of claim 12, wherein the regularly identifiable expression identifying operation is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language.

21. The apparatus of claim 12, wherein the input sequence of data symbols is representative of at least one of text data, transcribed spoken data, deoxyribonucleic acid sequence data, ribonucleic acid sequence data, amino-acid sequence data, audio sequence data, and video sequence data.

22. The apparatus of claim 12, wherein the input sequence of data symbols is representative of a voice mail message.

23. An article of manufacture for use in automatically processing an input sequence of data symbols, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

identifying at least one regularly identifiable expression in the input sequence of data symbols;

identifying at least a portion of information associated with the at least one regularly identifiable expression; and

extracting the portion of information.

24. Apparatus for automatically processing an input sequence of data symbols, the apparatus comprising:

a data capture device for obtaining the input sequence of data symbols;

at least one processor, operatively coupled to the data capture device, operative to:
(i) identify at least one regularly identifiable expression in the input sequence of data symbols; (ii) identify at least a portion of information associated with the at least one regular expression; and (iii) extract the portion of information;

5 memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations; and

 a data output device, operatively coupled to the at least one processor, for presenting the extracted portion of information to a user.

25. A method of automatically processing an input document, the method
10 comprising the steps of:

 identifying one or more regular expressions in the input document by comparing the input document to one or more previously-stored regular expressions to determine if there is a match between a portion of the input document and at least one of the previously-stored regular expressions; and

15 identifying at least a portion of information associated with the at least one regular expression for extraction.

**INFORMATION EXTRACTION FROM DOCUMENTS WITH REGULAR
EXPRESSION MATCHING**

Abstract of the Disclosure

Techniques are provided for enumerating regularly identifiable or stereotypical phrases that people commonly use to convey particular information, and where exactly in these phrases the particular information is to be found. In one embodiment, such phrases are referred to as “regular expressions.” Using such enumerated phrases, the invention is able to automatically identify them in an input data stream and then identify and extract the particular information associated with the phrase that is being sought, e.g., important or relevant information.

1500-211.APP

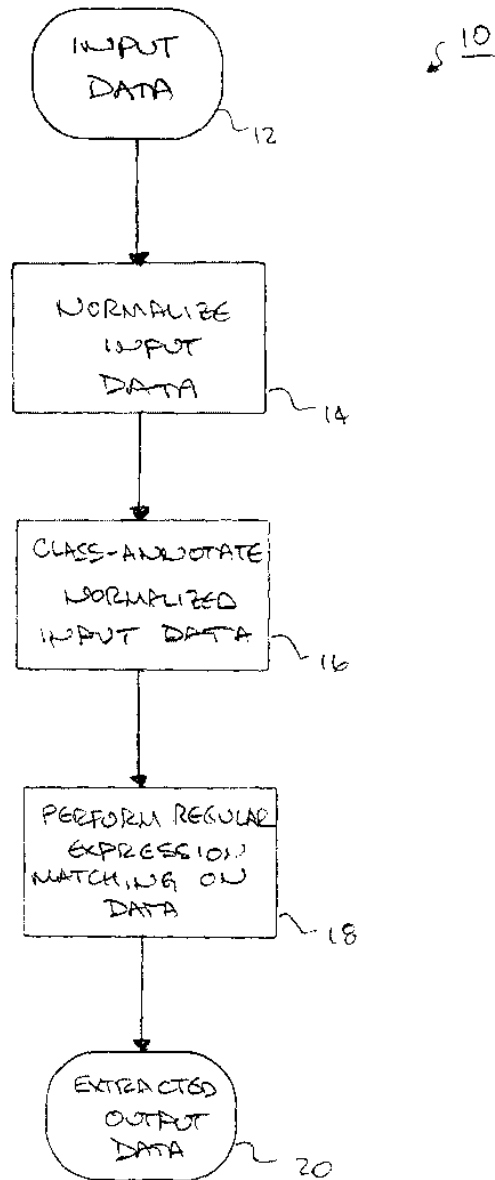


FIG. 1

2/2
Y02A20010253US1

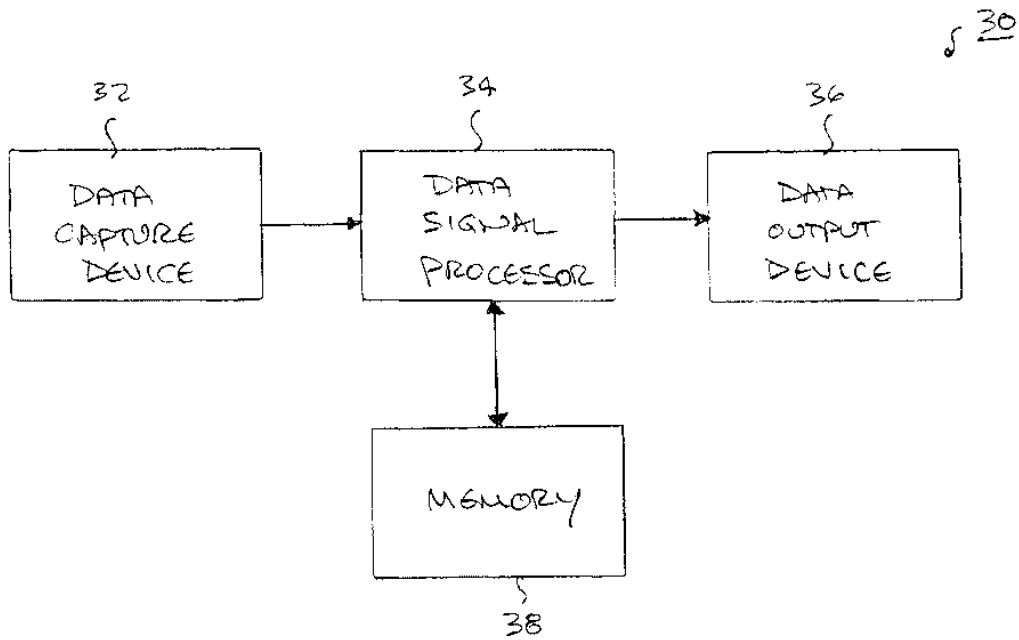


FIG. 2

DECLARATION

AS A BELOW NAMED INVENTOR, I hereby declare that:

My residence, post office address and citizenship are as stated next to my name.

I believe that I am the original, first and sole (*if only one name is listed below*), or an original, first and joint inventor (*if plural names are listed below*), of the subject matter which is claimed and for which a patent is sought on the invention entitled:
TITLE: INFORMATION EXTRACTION FROM DOCUMENTS WITH REGULAR EXPRESSION MATCHING

the specification of which is attached hereto or indicates an attorney docket no. YOR920010253US1, or:

☐ was filed in the U.S. Patent & Trademark Office on _____ and assigned Serial No. _____,

☐ and (*if applicable*) was amended on _____.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to patentability and to the examination of this application in accordance with Title 37, Code of Federal Regulations §1.56. I hereby claim foreign priority benefits under Title 35, U.S. Code §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate, or §365(a) of any PCT international application which designated at least one country other than the United States, or §119(e) of any United States provisional application(s), listed below and have also identified below any foreign applications for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

			Priority Claimed:
			Yes [] No []
(Application Number)	(Country)	(Day/Month/Year filed)	
_____	_____	_____	Yes [] No []
(Application Number)	(Country)	(Day/Month/Year filed)	
_____	_____	_____	Yes [] No []

I hereby claim the benefit under Title 35, U.S. Code §120, of any United States application(s), or §365(c), of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application(s) in the manner provided by the first paragraph of Title 35, U.S. Code §112, I acknowledge the duty to disclose information material to patentability as defined in Title 37, Code of Federal Regulations §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

(Application Serial Number)	(Filing Date)	(STATUS: patented, pending, abandoned)
_____	_____	_____
(Application Serial Number)	(Filing Date)	(STATUS: patented, pending, abandoned)
_____	_____	_____

I hereby appoint the following attorneys: **MANNY W. SCHECTER**, Reg. No. 31,722; **LAUREN C. BRUZZONE**, Reg. No. 35,082; **CHRISTOPHER A. HUGHES**, Reg. No. 26,914; **JOHN E. HOEL**, Reg. No. 26,279; **JOSEPH C. REDMOND, JR.**, Reg. No. 18,753; **STANLEY B. GREEN**, Reg. No. 24,351; **STEPHEN C. KAUFMAN**, Reg. No. 29,551; **ROBERT M. TREPP**, Reg. No. 25,933; **LOUIS P. HERZBERG**, Reg. No. 41,500; **DANIEL P. MORRIS**, Reg. No. 32,053; **PAUL J. OTTERSTEDT**, Reg. No. 37,411; **LOUIS J. PERCELLO**, Reg. No. 33,206; **DOUGLAS W. CAMERON**, Reg. No. 31,596; **MARIAN UNDERWEISER**, Reg. No. 46,134; **RICHARD M. LUDWIN**, Reg. No. 33,010; **MARC A. EHRlich**, Reg. No. 39,966; **ROBERT P. TASSINARI, JR.**, Reg. No. 36,030; **DEREK S. JENNINGS**, Reg. No. 41,473; **GAIL H. ZARICK**, Reg. No. 43,303; and **TIMOTHY M. FARRELL**, Reg. No. 37,321; each of them of **INTERNATIONAL BUSINESS MACHINES CORPORATION**, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598; to prosecute this application and to transact all business in the U.S. Patent and Trademark Office connected therewith and with any divisional, continuation, continuation-in-part, reissue or re-examination application, with full power of appointment and with full power to substitute an associate attorney or agent, and to receive all patents which may issue thereon, and request that all correspondence be addressed to:

William E. Lewis
 RYAN, MASON & LEWIS, LLP
 90 Forest Avenue
 Locust Valley, NY 11560
 Tel.: (516) 759-2946

I HEREBY DECLARE that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under §1001 of Title 18 U.S. Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

FULL NAME OF SOLE INVENTOR: Geoffrey G. Zweig Citizenship U.S.A.

Inventor's signature: _____ Date: _____

Residence & Post Office address: 30 Brookside Drive, Apt. 1A
Greenwich, Connecticut 06830

FULL NAME OF SECOND JOINT INVENTOR: Mukund Padmanabhan Citizenship India

Inventor's signature: _____ Date: _____

Residence & Post Office address: 19 Old Mamaroneck Road, #5E
White Plains, New York 10605

Figure 1. The effect of the number of trials on the number of correct responses. The number of correct responses was plotted against the number of trials for each condition. The number of correct responses increased with the number of trials for all conditions. The number of correct responses was highest for the condition with the highest number of trials (10 trials) and lowest for the condition with the lowest number of trials (2 trials).

07/30/20

Class	Subclass
ISSUE CLASSIFICATION	

PATENT NUMBER

U.S. UTILITY Patent Application

O.I.P.E.	PATENT DATE
 SCANNED SA(2) G.A. 	

APPLICATION NO.	CONT/PRIOR	CLASS	SUBCLASS	ART UNIT	EXAMINER
09/898289		710		2182	

CANTS Geoffrey Zweig
Mukund Padmanabhan

Information extraction from documents with regular expression matching

PTC-2040
12/99[illegible]

☐ TERMINAL DISCLAIMER	DRAWINGS			CLAIMS ALLOWED	
	Sheets Drwg.	Figs. Drwg.	Print Fig.	Total Claims	Print Claim for O.G.
☐ The term of this patent subsequent to _____ (date) has been disclaimed. ☐ The term of this patent shall not extend beyond the expiration date of U.S. Patent. No. _____ _____ _____	_____ (Assistant Examiner) (Date)			NOTICE OF ALLOWANCE MAILED	
	_____ (Primary Examiner) (Date)			ISSUE FEE	
				Amount Due	Date Paid
☐ The terminal _____ months of this patent have been disclaimed.	_____ (Legal Instruments Examiner) (Date)			ISSUE BATCH NUMBER	
WARNING: The information disclosed herein may be restricted. Unauthorized disclosure may be prohibited by the United States Code Title 35, Sections 122, 181 and 368. Possession outside the U.S. Patent & Trademark Office is restricted to authorized employees and contractors only.					

Form PTO-438A
(Rev. 8/99)

FILED WITH: ☐ DISK (CRF) ☐ FICHE ☐ CD-ROM
(Attached in pocket on right inside flap)

BEST AVAILABLE COPY

(FACE)

SEARCHED			
Class	Sub.	Date	Exmr.
710	65	7/28 -29	JK
710 705	65 16 1	11/23 -25/ 04 ↓	JK ↓

INTERFERENCE SEARCHED			
Class	Sub.	Date	Exmr.

SEARCH NOTES (INCLUDING SEARCH STRATEGY)		
	Date	Exmr.
EAST search	7/28 -29	JK
IEEE search	7/28	↓
EAST search	11/23-25/ 04	JK

Palo Alto Networks - Exhibit 1002
Palo Alto Networks v. Taasera - IPR2023-00443
Page 32 of 164

ISSUE SLIP STAPLE AREA (for additional cross references)

POSITION	INITIALS	ID NO.	DATE
FEE DETERMINATION	RH		7/9
O.I.P.E. CLASSIFIER		2	7-1701
FORMALITY REVIEW	TT	117	8/21/01
RESPONSE FORMALITY REVIEW	la	201	10-12-01

INDEX OF CLAIMS

✓ Rejected N Non-elected
 = ... Allowed I Interference
 - (Through numeral)... Canceled A Appeal
 + ... Restricted O Objected

Claim	Date
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	

Claim	Date
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

Claim	Date
101	
102	
103	
104	
105	
106	
107	
108	
109	
110	
111	
112	
113	
114	
115	
116	
117	
118	
119	
120	
121	
122	
123	
124	
125	
126	
127	
128	
129	
130	
131	
132	
133	
134	
135	
136	
137	
138	
139	
140	
141	
142	
143	
144	
145	
146	
147	
148	
149	
150	

BEST AVAILABLE COPY

If more than 150 claims or 10 actions
staple additional sheet here

(LEFT INSIDE)

10/2/01
505

07/03/01
01130 U.S. PTO

07-05-01

A

Attorney Docket No. YOR920010253US1

IN THE UNITED STATES
PATENT AND TRADEMARK OFFICE

PATENT APPLICATION

APPLICANT(S): Geoffrey G. Zweig
Mukund Padmanabhan

ASSIGNEE NAME: International Business
Machines Corporation

ASSIGNEE RESIDENCE: Armonk, New York

TITLE: Information Extraction From Documents
With Regular Expression Matching

ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231

SIR:

Enclosed are the following papers relating to the above-named application for patent:

Application - 18 pages of Specification, 5 pages of Claims and 1 page of Abstract
2 Sheets of informal drawing(s)
Declaration and Power of Attorney (unexecuted)

CLAIMS AS FILED				
	NO. FILED	NO. EXTRA	RATE	CALCULATIONS
Total Claims	25-20 =	5	x \$18 =	\$90
Independent Claims	5-3 =	2	x \$80 =	\$160
Multiple Dependent Claim(s), if applicable			\$270 =	\$0
Basic Fee				\$710
TOTAL FEE:				\$960

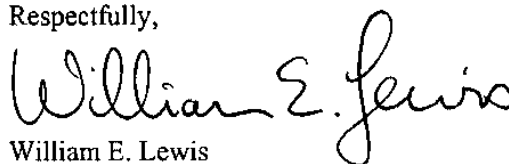
Please file the application and charge **International Business Machines Corporation Deposit Account No. 50-0510** the amount of \$960, to cover the filing fee. In the event of non-payment or improper payment of a required fee, the Commissioner is authorized to charge or to credit **International Business Machines Corporation Deposit Account No. 50-0510** as required to correct the error. Duplicate copies of this letter are enclosed.

Page 1 of 2

Attorney Docket No. YOR920010253US1

Please address all correspondence to: **Ryan, Mason & Lewis, LLP, 90 Forest Avenue, Locust Valley, New York 11560**. Telephone calls should be made to the under-signed attorney at (516) 759-2946.

Respectfully,



Date: July 3, 2001

William E. Lewis
Reg. No. 39,274
Attorney for Applicant(s)
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, New York 11560

REG. NO. 39,274

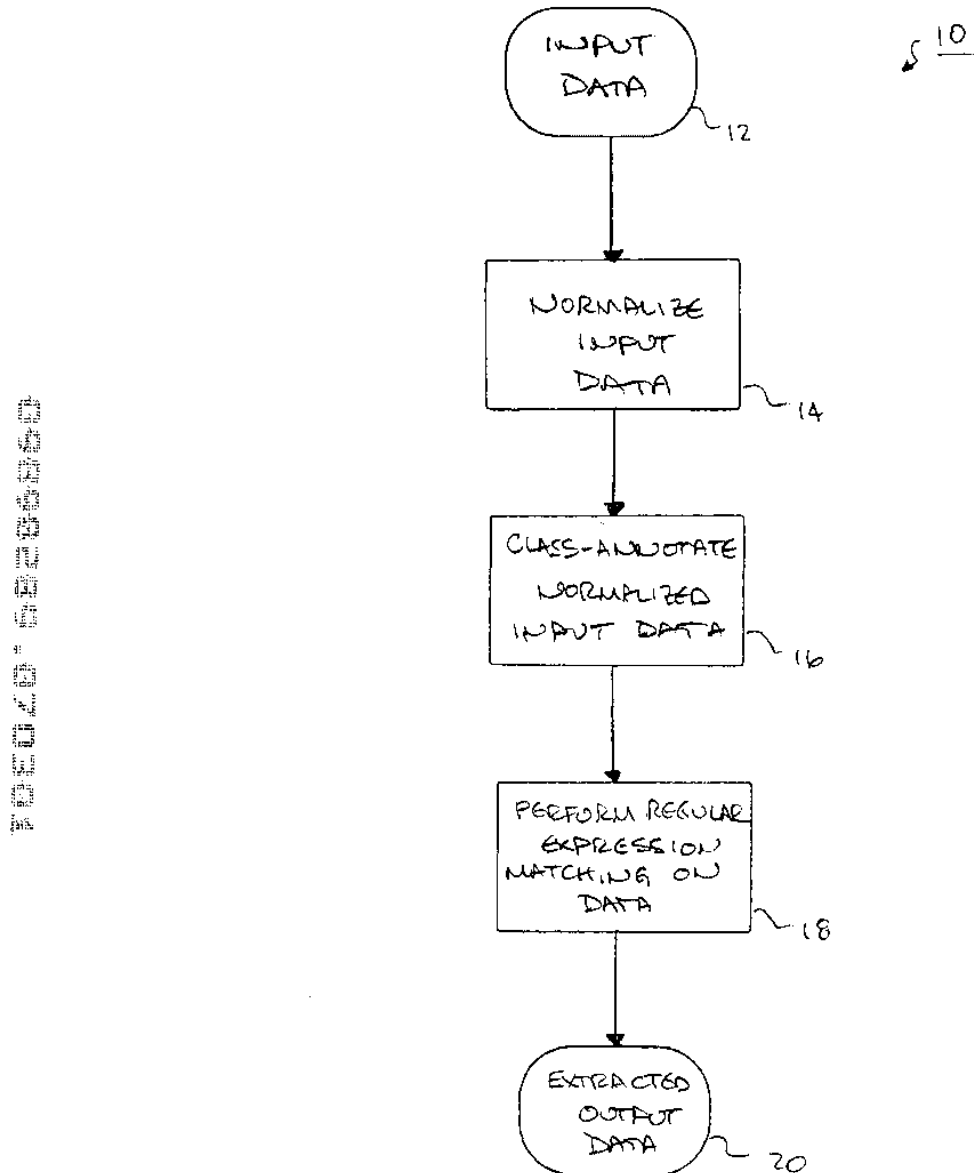


FIG. 1

2/2

402920010253US1

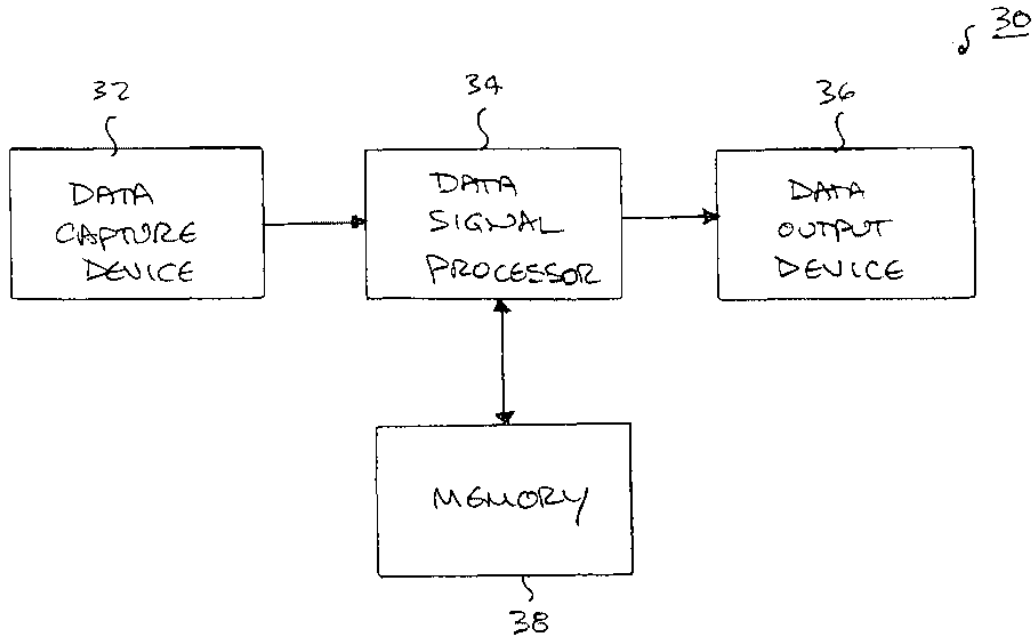


FIG. 2

**INFORMATION EXTRACTION FROM DOCUMENTS WITH REGULAR
EXPRESSION MATCHING**

Field of the Invention

5 The present invention relates to data processing techniques and, more particularly,
to techniques for providing automatic information extraction from data streams such as
text and spoken documents.

Background of the Invention

10 There are many situations in which it is desirable to extract key pieces of
information from documents. For example, in transcribed voice mail messages, the name
of the caller and any return numbers that were left are crucial for summarizing the call.
Or when resumes are submitted to a company along with a cover letter, it is desirable to
extract the job objective and salary requirements of the applicant, in order to determine if
a suitable match exists. This invention comprises a simple, efficient, and effective way of
extracting such information.

15 In the past, several techniques have been developed to solve the somewhat
simpler problem of "named entity extraction." In this task, the goal is to identify all
occurrences of certain classes of words in a document. For example, all the
person-names, city-names, dates, and times might be identified. One way of identifying
such entities is to train a statistical classification system to tag each word in the document
20 as either a "person-name," "city-name," "date," "time," or "other" word. Examples of
this sort of approach are disclosed in, e.g., Ratnaprkhi, "A Maximum Entropy Part of
Speech Tagger," Conference on Empirical Methods in Natural Language Processing,
University of Pennsylvania, 1996; and Brill, "Transformation-Based Error-Driven
Learning and Natural Language Processing: A Case Study in Part of Speech Tagging,"
25 Computational Linguistics, December 1995, the disclosures of which are incorporated by
reference herein. While the problem of extracting key pieces of information can also be

viewed as a tagging problem, in which each word is tagged as either “key information” or “irrelevant,” such a tagging approach fails to explicitly identify the portions of text that are important.

Thus, there is a need for data processing techniques which explicitly identify portions of data that are sought to be identified, rather than only implicitly identifying, or tagging, such portions of data.

Summary of the Invention

The present invention provides techniques for exploiting the readily-identifiable structure of language to explicitly identify portions of data in a document that a user seeks to be identified, e.g., relevant or important information. It is to be understood that the term “document,” as used herein in accordance with the invention, is intended to broadly refer to a sequence of symbols (e.g., letters, numbers, characters, etc.), such as a data stream representative of text and/or spoken messages. However, it is to be understood that the invention applies to many input forms other than typed text or spoken messages. By way of further example only, the invention may also be applied to information extraction from data streams representing DNA (deoxyribonucleic acid) sequences, RNA (ribonucleic acid) sequences, amino-acid sequences, and audio and video sequences that are preprocessed into a discrete symbolic form.

More specifically, the present invention realizes that, in common linguistic usage, people typically convey information in highly stereotyped ways. So, for example, in a phone call, the caller is likely to identify himself in one of just a few ways, e.g., “Hi <recipient-name>, it’s <caller-name>” (e.g., “Hi John, it’s Bob”) or “<recipient-name>, <caller-name> here” (e.g., “John, Bob here”). Or in a cover letter, a job applicant is likely to express his interest with a phrase like “I am looking for a job in <field>” (e.g., “I am looking for a job in electrical engineering”).

In accordance with this realization, the present invention provides techniques for enumerating regularly identifiable or stereotypical phrases that people commonly use to

convey particular information, and where exactly in these phrases the particular information is to be found. In one embodiment, such phrases are referred to as "regular expressions." Using such enumerated phrases, the invention is able to automatically identify them in an input data stream and then identify and extract the particular information associated with the phrase that is being sought, e.g., important or relevant information.

This approach has several advantages over the previously mentioned tagging approach. First, the phrases are readily expressible as regular expressions, and this allows them to be stored in one of several common programming languages, e.g., as a "flex" or "perl" program. Both these languages have built-in support for regular expressions. Secondly, the use of regular-expression technology enables extremely fast pattern matching and information extraction with highly optimized standard programs. Thirdly, the phrases can be identified without the expensive and time consuming step of gathering and annotating a large "training" database. Finally, if desired, the technique can be combined with statistical based procedures, e.g., by using the phrases as features in a statistical system, or by using a statistical procedure to automatically collect a large pool of candidate phrases which can then be selected for inclusion by a human.

One skilled in the art will realize various applications for the invention. By way of example only, such enumerated phrases may be automatically identified in a transcribed voice mail message, and the important information in the message, e.g., a phone number or caller's name, may be automatically extracted and presented to the user or stored for subsequent use.

These and other objects, features and advantages of the present invention will become apparent from the following detailed description of illustrative embodiments thereof, which is to be read in connection with the accompanying drawings.

Brief Description of the Drawings

FIG. 1 is a flow diagram illustrating an information extraction methodology according to an embodiment of the present invention; and

FIG. 2 is a block diagram illustrating a generalized hardware architecture of a data processing system suitable for implementing an information extraction methodology according to the present invention.

Detailed Description of Preferred Embodiments

As will be explained in detail below, the present invention provides techniques for using stereotypical phrases or “regular expressions” to identify the information-bearing portions of a document, and thereby explicitly identify the information found therein. Regular expressions, as have been discussed in Aho et al., “Compilers: Principles, Techniques, and Tools,” Addison-Wesley, 1985, the disclosure of which is incorporated by reference herein, comprise symbolic strings, typically word sequences, that can be constructed by the basic operations of “and” and “or” acting on an atomic alphabet of symbols. So, for example, the pattern “the dog” produces a match only when, in the data being considered, the symbol “the” is followed by the symbol “dog.” This represents the operation of concatenation or “and.” The pattern “the (dog | cat)” illustrates the “or” operation, i.e., the pattern matches either the string “the dog” or “the cat.” The operation of “classing” is also possible. So, for example, a “pet” class may be defined as either “dog” or “cat.”

`<animal> ::= (dog | cat)`

Then, subsequent regular expressions can make use of this class, for example, the pattern: “the <pet>.” A wide variety of extensions to regular expressions exist, for example, restrictions on where in a document a pattern is found, or restrictions on the length of repeated sequences in a regular expression. In accordance with an illustrative embodiment of the invention, the term “regular expression” is taken to mean any form of pattern that is matchable in the *flex*, *lex*, or *perl* programming languages, although the

In addition to being used as a stand-alone method for information extraction, regular expressions can be in combination with statistical techniques. For example, the existence of a pattern such as "Hi <person-name>, it's <person-name>" can be used as a feature in a maximum-entropy word-tagging system.

5 Referring now to FIG. 1, a flow diagram illustrates an information extraction methodology according to an embodiment of the present invention. As shown in FIG. 1, the methodology 10 begins by obtaining the input data 12 to be processed. As mentioned above, the input data is a sequence of discrete symbols or a document, such as may be representative of text and/or a transcribed spoken message.

10 In step 14, the input data is normalized. This process may include standardizing any characteristics of the input that are irrelevant to, or a hindrance to, information extraction. An example of normalization, for text input, is to change all the letters in the text stream to lower-case and to replace multiple blank spaces between words by a single blank space. This makes subsequent pattern matching easier.

15 By way of example, the following is a sample *flex* program for normalizing text input:

%{

#include <iostream.h>
#include <ctype.h>

20 %}

%option noyywrap
%option never-interactive

ws [t]+

%%

25 - {cout << " ";}
[!\\.,\"'() ?] {;}

YOR920010253US1

```

\<[^\>]*\>{ws}*      {}
{ws}                  {cout<<" ";}
                      {cout<<char(tolower(*yytext));}

```

```
%%
```

```

5  main( )
    {
        yylex();
    }

```

In step 16, one or more previously-stored word (or, in general, one or more symbols) lists are used to identify and mark occurrences of words that belong to certain classes. By way of example, such a class-annotation operation may include marking all proper names by prefixing them with an "!" mark.

By way of example, the following is a sample C++ program for annotating names identified in a normalized text stream:

```

15  #include <iostream.h>
    #include <fstream.h>
    #include <stdlib.h>
    #include <set>
    #include <string>

20  main()
    {
        set<string>name;
        ifstream namesin("../data/person_names"); // a text file listing names
        if (!namesin)
25      {
            cout << "Unable to find ../data/person_names\n";
            exit(1);
        }

        while (!namesin.eof())
30      {
            string s;

```

```

namesin >> s >> ws;
name.insert(s);
}
namesin.close();
5   while (!cin.eof())
    {
        string s;
        cin >> s >> ws;
        if (name.count(s))
10      cout << "!" << s << " ";
        else
            cout << s << " ";
    }
}

```

15 In step 18, the normalized and class-annotated text is read by a regular-expression matching program, and the information-bearing portions of the text are analyzed and extracted. In a preferred embodiment, the regular expressions are specified in and matched by a specialized regular expression matching program, such as may be implemented in *flex*. In this process, it is understood that all the mechanisms available in

20 such a programming language may be used. These include the use of special start states, backtracking, context-sensitive matching, and embedded special-purpose "C" code.

Specifically, the matching operation in step 18 takes the normalized, class-annotated input data and identifies the regular expressions in the data by comparing the data to previously-stored regular expressions. These previously-stored regular

25 expressions may include, for example, trigger prefixes, trigger suffixes, word classes, and other matchable patterns. When a match occurs, the structure of the regular expression is analyzed to identify the information-bearing portion of the expression. The information is then extracted from the expression.

By way of example, the following is a sample *flex* program for identifying caller

30 names and phone numbers in a transcribed voice mail messages. As is evident, the program enumerates a plurality of regular expressions, categorized as

YOR920010253US1

PHONE_NUMBER_STATE, NAME_STATE, EXTENDED_NAME_STATE, and
TERMINATION_STATE, that are considered during the matching step 18:

```
%{  
#include <iostream.h>  
5 #include <string.h>  
  
const char *number_type;  
const char *tl = "tieline";  
const char *extension = "extension";  
const char *pager = "pager";  
10 const char *fax = "fax number";  
const char *none = "";  
  
char number_buff[10000];  
  
%}  
  
%option noyywrap  
15 %option never-interactive  
  
/* many patterns like IN-GENERIC_NAME-AT-DIGIT+unaccounted for */  
  
spelled_digit  
((zero|oh|one|two|three|four|five|six|seven|eight|nine|hundred|thousand|ten|  
eleven|twelve|thirteen|fourteen|fifteen|sixteen|seventeen|eighteen|nineteen|twenty|  
20 thirty|forty|fifty|sixty|seventy|eighty|ninety|hundred)[ ]?)  
numeric_digit ((1|2|3|4|5|6|7|8|9|10){ }?)  
digit ([ ]?{spelled_digit}|[ ]?{numeric_digit})  
phone_number ((country[ ]code[ ]{digit}{2,})?(toll[ ]free)?(area[ ]|area[ ]code[  
])?{digit}{2,} ((or[ ]|that's[ ]|sorry[ ])|{digit}{2,})?(option[ ]{digit}+)?(extension[  
25 ]{digit}{2,})?)? tie_line ([ ]?(tie[ ]line|tieline)([ ]number)?)  
name !([a-z']*)+  
  
%x PHONE_NUMBER_STATE  
%x NAME_STATE  
%x EXTENDED_NAME_STATE  
30 %x TERMINATION_STATE
```

YOR920010253US1

9

```

%%

(1|2|3|4|5|6|7|8|9|0)+ ":" {cout << yytext << endl; /* get header */}

{digit}({digit}|"extension"){3,} {strcpy(number_buff, yytext); BEGIN
TERMINATION_STATE;}

5  <TERMINATION_STATE>"thank you" |
  <TERMINATION_STATE>"thanks" |
  <TERMINATION_STATE>"take care" |
  <TERMINATION_STATE>"bye" |
  <TERMINATION_STATE>"talk to you" |
10 <TERMINATION_STATE>"okay bye" |
  <TERMINATION_STATE>"have a "(good | great)" day" {cout << "Number: "
  << number_buff << endl; BEGIN INITIAL;}
  <TERMINATION_STATE><<EOF>> {cout << "Number: "number_buff << endl;
  BEGIN INITIAL;}
15 <TERMINATION_STATE>.{unput(*yytext); BEGIN INITIAL;}

  "on extension" |
  "at extension" |
  "extension" |
  "extension is" {number_type = extension; BEGIN PHONE_NUMBER_STATE;}

20 "my "{tie_line} "is" |
  "my "{tie_line}" here is" |
  {tie_line}"is" |
  "{tie_line}" " {number_type = tl; BEGIN PHONE_NUMBER_STATE;}

  "again is" |
25 "again" |
  "back" |
  "call me" |
  "call me at" |
  "call me back at" |
30 "call me "[a-zA-Z]+ "at" |
  "call me back on" |
  "call me on" |
  "call back at" |
  "call back on" |
35 "give me a buzz" |

```

YOR920010253US1

10

"give me buzz at" |
 "give me a holler" |
 "give me a holler at" |
 "give me a ring" |
 5 "give me a call" |
 "give me a ring at" |
 "give me a call at" |
 "give me a call at home" |
 "give me a call at work" |
 10 "give me a call at your convenience" |
 "give me a call back at" |
 "give me a call back on" |
 "give me a call back with that information at" |
 "give me a call if you get a chance" |
 15 "give me a call please" |
 "give me a call please at" |
 "give me a call when you get a chance" |
 "give me a call when you can" |
 "give me a call when you can please" |
 20 "give us a call" |
 "give us a call at" |
 "here" |
 "i can be reached at" |
 "i am at" |
 25 "i'm at" |
 "i'll be at" |
 "i'm on" |
 "leave me a message" |
 "leave me a message at" |
 30 "my number is" |
 "my number here is" |
 "my number here in "(!?[a-z]+[])"is" |
 "my number's" |
 "my phone number's" |
 35 "number again is" |
 "number at home is" |
 "number at work is" |
 "number here is" |
 "number is" |
 40 "number's" |

11


```

c=strchr(c+1, ' '); cout << "Name: " << (c+1) << endl;}

"good morning "{name}" " |
"yeah"{name}" " |
({name}[ ])? "this is" |
5 {name}"it's" |
{name}"hi" |
"hi i'm" |
"my name is" |
"my name's" |
10 "the name is"
"the name's" |
"hey"({name}[ ])? "this"("is")? |
"hi"({name}[ ])? "this"("is")? |
({name}[ ])? "hi it's" |
15 ({name}[ ])? "hey it's" |
({name}[ ])? "how you doing it's" |
{name}"this" {BEGIN NAME_STATE;}

<NAME_STATE>({name}[ ])+ {cout << "Name: " << yytext << endl;
number_type=none;
20 BEGIN PHONE_NUMBER_STATE;}

<NAME_STATE>("me"|"mom"|"your sis"|"your sister"|"your wife"|"your brother"|"
"your father"|"your husband") {cout << "Name: " << yytext << endl; BEGIN
INITIAL;}

<NAME_STATE>({name}[ ])+"from" |
25 <NAME_STATE>({name}[ ])+"calling from" |
<NAME_STATE>({name}[ ])+"i'm calling from" |
<NAME_STATE>({name}[ ])+"with" |
<NAME_STATE>({name}[ ])+"i'm with" |
<NAME_STATE>({name}[ ])+"at" |
30 <NAME_STATE>({name}[ ])+"over at" |
<NAME_STATE>({name}[ ])+"down at" |
<NAME_STATE>({name}[ ])+"over in" |
<NAME_STATE>({name}[ ])+"up in" |
<NAME_STATE>({name}[ ])+"down in" |
35 <NAME_STATE>({name}[ ])+"up on" |
<NAME_STATE>({name}[ ])+"down on" |

```



```

5  <EXTENDED_NAME_STATE>"about"|
    <EXTENDED_NAME_STATE>"again"|
    <EXTENDED_NAME_STATE>"wanting"|
    <EXTENDED_NAME_STATE>"appreciate"|
    <EXTENDED_NAME_STATE>"was"|
    <EXTENDED_NAME_STATE>"we're"|
    <EXTENDED_NAME_STATE>"or"|
    <EXTENDED_NAME_STATE>"you've"|
    <EXTENDED_NAME_STATE>"right"|
10  <EXTENDED_NAME_STATE>"first"|
    <EXTENDED_NAME_STATE>[ ]?"tieline"|
    <EXTENDED_NAME_STATE>[ ]?"extension"|
    <EXTENDED_NAME_STATE>" " {digit}|
    <EXTENDED_NAME_STATE>"please" {cout << endl; char *q=yytext+strlen(yytext);
15  while (q>yytext) unput(*--q); number_type=none; BEGIN
    PHONE_NUMBER_STATE;}
    <EXTENDED_NAME_STATE>. {ECHO;}}

    .{;}

    %%

20  main()
    {
        yylex();
        cout << endl;
    }

```

25 Thus, by way of example, if the voice message input to the methodology 10 contains the phrase "call me back at 555-1234," the above program determines that the phrase "call me back at" matches one of the enumerated expressions. Then, the information associated with the phrase "call me back at," namely, "555-1234," is identified as the information-bearing portion of the expression, and extracted.

30 The extracted output data 20 may then be presented to a user. For example, this may include displaying the information to the user on a display or converting the information to audio and audibly presenting the information to the user in accordance

with an output speaker. The extracted data may also be used to take some other specified action, e.g., place extracted data in an electronic phone book, notify user of a message from a particularly important person, notify a user of a message on a particular topic, etc.

Referring now to FIG. 2, a block diagram illustrates a generalized hardware architecture of a data processing system suitable for implementing an information extraction methodology according to the present invention. As shown, the system 30 includes a data capture device 32, a data signal processor 34, a data output device 36, and memory 38.

The data capture device 32 generally obtains the data stream that is to be processed by the system 30. The device's specific structure and operations depend on the application in which the system is employed. For example, in a voice mail application, the data capture device may be a speech processing device capable of taking an audio stream recorded in accordance with a voice mail system and converting or transcribing the audio stream to a discrete data or symbol sequence. The device may, for example, utilize conventional speech recognition techniques or some other known transcription techniques. In a text based application, the device may be a digital scanner capable of converting a text document into a discrete data or symbol sequence. The device may, for example, utilize conventional optical character recognition techniques. Of course, depending on the data type, the data capture device may take on alternate forms.

The data signal processor 34 generally performs and/or controls the operations of the methodologies of the present invention, in conjunction with the memory 38. That is, the signal processor performs and/or controls the steps depicted in the flow diagram of FIG. 1.

It is to be appreciated that the term "processor" as used herein is intended to include any processing device, such as, for example, one that includes a CPU (central processing unit) or other processing circuitry. For example, the processor may be a digital signal processor, as is known in the art. Also the term "processor" may refer to one or more individual processors. The term "memory" as used herein is intended to

include any memory devices associated with a processor or CPU, such as, for example, RAM, ROM, a fixed memory device (e.g., hard drive), a removable memory device (e.g., diskette), flash memory, etc. It is to be understood that the previously-generated word classes and regular expressions used in the class-annotation and expression matching operations, described above in accordance with the invention, may be stored in memory 38 and accessed when required.

Accordingly, one or more computer software programs including instructions or code for performing the methodologies of the invention, as described herein, may be stored in one or more of the associated memory devices (e.g., ROM, fixed or removable memory) and, when ready to be utilized, loaded in part or in whole (e.g., into RAM) and executed by a CPU.

The data output device 36 generally presents the data extracted from the input data stream to a user. Again, the device's specific structure and operations depend on the application in which the system is employed. For example, in a voice mail application, the data capture device may be an audio output system capable of taking the extracted discrete data and converting the data to a form to be audibly output to the user. The device may, for example, utilize conventional text-to-speech or other speech synthesis techniques. Such techniques may also be used in a text based application. Alternatively, the device may be a display for visually presenting the extracted data to the user. Of course, depending on the data type, the data output device may take on alternate forms.

It is to be understood that the elements illustrated in the figures may be implemented in various forms of hardware, software, or combinations thereof, e.g., one or more digital signal processors with associated memory, application specific integrated circuit(s), functional circuitry, one or more appropriately programmed general purpose digital computers with associated memory, etc. For example, the information extraction methodologies of the invention may be implemented in various personal computing devices. Given the teachings of the invention provided herein, one of ordinary skill in the

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2
--	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	---

Claims

What is claimed is:

1. A method of automatically processing an input sequence of data symbols, the method comprising the steps of:

identifying at least one regularly identifiable expression in the input sequence of data symbols;

identifying at least a portion of information associated with the at least one regularly identifiable expression; and

extracting the portion of information.

2. The method of claim 1, wherein the regularly identifiable expression identifying step comprises comparing the input sequence of data symbols to one or more previously-stored regularly identifiable expressions to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions.

3. The method of claim 1, further comprising the step of normalizing the input sequence of data symbols prior to identifying the regularly identifiable expression.

4. The method of claim 1, further comprising the step of identifying one or more classes of data symbols in the input sequence prior to identifying the regularly identifiable expression.

5. The method of claim 1, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically precedes a particular portion of information.

YOR920010253US1

19

6. The method of claim 1, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically follows a particular portion of information.

7. The method of claim 1, wherein the extracted portion of information is used to take a specified action.

8. The method of claim 1, wherein the extracted portion of information is at least one of visually and audibly presented to the user.

9. The method of claim 1, wherein the regularly identifiable expression identifying step is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language.

10. The method of claim 1, wherein the input sequence of data symbols is representative of at least one of text data, transcribed spoken data, deoxyribonucleic acid sequence data, ribonucleic acid sequence data, amino-acid sequence data, audio sequence data, and video sequence data.

11. The method of claim 1, wherein the input sequence of data symbols is representative of a voice mail message.

12. Apparatus for automatically processing an input sequence of data symbols, the apparatus comprising:

at least one processor operative to: (i) identify at least one regularly identifiable expression in the input sequence of data symbols; (ii) identify at least a portion of information associated with the at least one regularly identifiable expression; and (iii) extract the portion of information; and

YOR920010253US1

20

memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations.

13. The apparatus of claim 12, wherein the regularly identifiable expression identifying operation comprises comparing the input sequence of data symbols to one or more regularly identifiable expressions, previously stored in the memory, to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions.

14. The apparatus of claim 12, wherein the at least one processor is further operative to normalize the input sequence of data symbols prior to identifying the regularly identifiable expression.

15. The apparatus of claim 12, wherein the at least one processor is further operative to identify one or more classes of data symbols in the input sequence prior to identifying the regularly identifiable expression.

16. The apparatus of claim 12, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically precedes a particular portion of information.

17. The apparatus of claim 12, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically follows a particular portion of information.

18. The apparatus of claim 12, wherein the extracted portion of information is used to take a specified action.

19. The apparatus of claim 12, wherein the extracted portion of information is at least one of visually and audibly presented to the user.

20. The apparatus of claim 12, wherein the regularly identifiable expression identifying operation is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language.

21. The apparatus of claim 12, wherein the input sequence of data symbols is representative of at least one of text data, transcribed spoken data, deoxyribonucleic acid sequence data, ribonucleic acid sequence data, amino-acid sequence data, audio sequence data, and video sequence data.

22. The apparatus of claim 12, wherein the input sequence of data symbols is representative of a voice mail message.

23. An article of manufacture for use in automatically processing an input sequence of data symbols, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

identifying at least one regularly identifiable expression in the input sequence of data symbols;

identifying at least a portion of information associated with the at least one regularly identifiable expression; and

extracting the portion of information.

24. Apparatus for automatically processing an input sequence of data symbols, the apparatus comprising:

a data capture device for obtaining the input sequence of data symbols;

at least one processor, operatively coupled to the data capture device, operative to:
(i) identify at least one regularly identifiable expression in the input sequence of data
symbols; (ii) identify at least a portion of information associated with the at least one
regular expression; and (iii) extract the portion of information;

5 memory, operatively coupled to the at least one processor, for storing at least a
portion of results associated with the identifying and extracting operations; and

a data output device, operatively coupled to the at least one processor, for
presenting the extracted portion of information to a user.

25. A method of automatically processing an input document, the method
10 comprising the steps of:

identifying one or more regular expressions in the input document by comparing
the input document to one or more previously-stored regular expressions to determine if
there is a match between a portion of the input document and at least one of the
previously-stored regular expressions; and

15 identifying at least a portion of information associated with the at least one regular
expression for extraction.

**INFORMATION EXTRACTION FROM DOCUMENTS WITH REGULAR
EXPRESSION MATCHING**

Abstract of the Disclosure

Techniques are provided for enumerating regularly identifiable or stereotypical phrases that people commonly use to convey particular information, and where exactly in these phrases the particular information is to be found. In one embodiment, such phrases are referred to as "regular expressions." Using such enumerated phrases, the invention is able to automatically identify them in an input data stream and then identify and extract the particular information associated with the phrase that is being sought, e.g., important or relevant information.

1500-211.APP

DECLARATION

AS A BELOW NAMED INVENTOR, I hereby declare that:

My residence, post office address and citizenship are as stated next to my name.

I believe that I am the original, first and sole (*if only one name is listed below*), or an original, first and joint inventor (*if plural names are listed below*), of the subject matter which is claimed and for which a patent is sought on the invention entitled:

TITLE: INFORMATION EXTRACTION FROM DOCUMENTS WITH REGULAR EXPRESSION MATCHING

the specification of which is attached hereto or indicates an attorney docket no. YOR920010253US1, or:

☐ was filed in the U.S. Patent & Trademark Office on _____ and assigned Serial No. _____,

☐ and (*if applicable*) was amended on _____.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to patentability and to the examination of this application in accordance with Title 37, Code of Federal Regulations §1.56. I hereby claim foreign priority benefits under Title 35, U.S. Code §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate, or §365(a) of any PCT international application which designated at least one country other than the United States, or §119(e) of any United States provisional application(s), listed below and have also identified below any foreign applications for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Priority Claimed:

Yes [] No []

(Application Number) (Country) (Day/Month/Year filed)

Yes [] No []

(Application Number) (Country) (Day/Month/Year filed)

I hereby claim the benefit under Title 35, U.S. Code §120, of any United States application(s), or §365(c), of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International applications(s) in the manner provided by the first paragraph of Title 35, U.S. Code §112, I acknowledge the duty to disclose information material to patentability as defined in Title 37, Code of Federal Regulations §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

(Application Serial Number) (Filing Date) (STATUS: patented, pending, abandoned)

(Application Serial Number) (Filing Date) (STATUS: patented, pending, abandoned)

I hereby appoint the following attorneys: MANNY W. SCHECTER, Reg. No. 31,722; LAUREN C. BRUZZONE, Reg. No. 35,082; CHRISTOPHER A. HUGHES, Reg. No. 26,914; JOHN E. HOEL, Reg. No. 26,279; JOSEPH C. REDMOND, JR., Reg. No. 18,753; STANLEY B. GREEN, Reg. No. 24,351; STEPHEN C. KAUFMAN, Reg. No. 29,551; ROBERT M. TREPP, Reg. No. 25,933; LOUIS P. HERZBERG, Reg. No. 41,500; DANIEL P. MORRIS, Reg. No. 32,053; PAUL J. OTTERSTEDT, Reg. No. 37,411; LOUIS J. PERCELLO, Reg. No. 33,206; DOUGLAS W. CAMERON, Reg. No. 31,596; MARIAN UNDERWEISER, Reg. No. 46,134; RICHARD M. LUDWIN, Reg. No. 33,010; MARC A. EHRLICH, Reg. No. 39,966; ROBERT P. TASSINARI, JR., Reg. No. 36,030; DEREK S. JENNINGS, Reg. No. 41,473; GAIL H. ZARICK, Reg. No. 43,303; and TIMOTHY M. FARRELL, Reg. No. 37,321; each of them of INTERNATIONAL BUSINESS MACHINES CORPORATION, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598; to prosecute this application and to transact all business in the U.S. Patent and Trademark Office connected therewith and with any divisional, continuation, continuation-in-part, reissue or re-examination application, with full power of appointment and with full power to substitute an associate attorney or agent, and to receive all patents which may issue thereon, and request that all correspondence be addressed to:

William E. Lewis
 RYAN, MASON & LEWIS, LLP
 90 Forest Avenue
 Locust Valley, NY 11560
 Tel.: (516) 759-2946

I HEREBY DECLARE that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under §1001 of Title 18 U.S. Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

FULL NAME OF SOLE INVENTOR: Geoffrey G. Zweig Citizenship U.S.A.

Inventor's signature: _____ Date: _____

Residence & Post Office address: 30 Brookside Drive, Apt. 1A
Greenwich, Connecticut 06830

FULL NAME OF SECOND JOINT INVENTOR: Mukund Padmanabhan Citizenship India

Inventor's signature: _____ Date: _____

Residence & Post Office address: 19 Old Mamaroneck Road, #5E
White Plains, New York 10605

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT(S): Geoffrey G. Zweig et al.
SERIAL NO.: 09/898,289
FILED: July 3, 2001
FOR: INFORMATION EXTRACTION FROM DOCUMENTS
WITH REGULAR EXPRESSION MATCHING

ASSOCIATE POWER OF ATTORNEY

Please recognize **JOSEPH B. RYAN**, Reg. No. 37,922; **KEVIN M. MASON**, Reg. No. 36,597; **WILLIAM E. LEWIS**, Reg. No. 39,274; **ROBERT J. MAURI**, Reg. No. 41,180; **WAYNE L. ELLENBOGEN**, Reg. No. 43,602; and **JAMES M. LOEFFLER**, Reg. No. 37,873; each of them of **RYAN, MASON & LEWIS, LLP**, 90 Forest Avenue, Locust Valley, New York 11560 as associate attorneys in the above-mentioned application, with full power to prosecute said application, to make alterations and amendments therein, and to transact all business in the Patent and Trademark Office connected therewith.

Telephone calls should be made to William E. Lewis by dialing (516) 759-2946.

All written communications are to be sent to **Ryan, Mason & Lewis, LLP**, 90 Forest Avenue, Locust Valley, New York 11560.

Dated: OCT. 1, 2001



Paul J. Offerstedt
Registration No. 37,411
Attorney for Applicant(s)

International Business Machines
Corporation
T.J. Watson Research Center
Route 134 and Kitchawan Road
Yorktown Heights, New York 10598

PATENT APPLICATION SERIAL NO. _____

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

07/10/2001 RHARIS1 00000002 500510 09898289

01 FC:101	710.00 CH
02 FC:102	160.00 CH
03 FC:103	90.00 CH

PTO-1556
(5/87)

*U.S. GPO: 2000-468-987/39595

PATENT APPLICATION FEE DETERMINATION RECORD
Effective October 1, 2000

Application or Docket Number

09898289

CLAIMS AS FILED - PART I

	(Column 1)	(Column 2)
TOTAL CLAIMS	25	
FOR	NUMBER FILED	NUMBER EXTRA
TOTAL CHARGEABLE CLAIMS	25 minus 20 = *	5
INDEPENDENT CLAIMS	5 minus 3 = *	2
MULTIPLE DEPENDENT CLAIM PRESENT ☐		

* If the difference in column 1 is less than zero, enter "0" in column 2

SMALL ENTITY
TYPE ☐ OR

OTHER THAN SMALL ENTITY

RATE	FEE	OR	RATE	FEE
BASIC FEE	355.00	OR	BASIC FEE	710.00
X\$ 9=		OR	X\$18=	90
X40=		OR	X80=	160
+135=		OR	+270=	
TOTAL		OR	TOTAL	960

CLAIMS AS AMENDED - PART II

	(Column 1)	(Column 2)	(Column 3)
AMENDMENT A	CLAIMS REMAINING AFTER AMENDMENT	HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA
Total	*	Minus	**
Independent	*	Minus	***
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM ☐			

SMALL ENTITY OR **OTHER THAN SMALL ENTITY**

RATE	ADDITIONAL FEE	OR	RATE	ADDITIONAL FEE
X\$ 9=		OR	X\$18=	
X40=		OR	X80=	
+135=		OR	+270=	
TOTAL ADDIT. FEE		OR	TOTAL ADDIT. FEE	

	(Column 1)	(Column 2)	(Column 3)
AMENDMENT B	CLAIMS REMAINING AFTER AMENDMENT	HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA
Total	*	Minus	**
Independent	*	Minus	***
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM ☐			

RATE	ADDITIONAL FEE	OR	RATE	ADDITIONAL FEE
X\$ 9=		OR	X\$18=	
X40=		OR	X80=	
+135=		OR	+270=	
TOTAL ADDIT. FEE		OR	TOTAL ADDIT. FEE	

	(Column 1)	(Column 2)	(Column 3)
AMENDMENT C	CLAIMS REMAINING AFTER AMENDMENT	HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA
Total	*	Minus	**
Independent	*	Minus	***
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM ☐			

RATE	ADDITIONAL FEE	OR	RATE	ADDITIONAL FEE
X\$ 9=		OR	X\$18=	
X40=		OR	X80=	
+135=		OR	+270=	
TOTAL ADDIT. FEE		OR	TOTAL ADDIT. FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20."

***If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3."

The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

CLAIMS ONLY							SERIAL NO. 09898289	FILING DATE					
							APPLICANT(S)						
CLAIMS													
	AS FILED		AFTER 1st AMENDMENT		AFTER 2nd AMENDMENT			*		*		*	
	IND.	DEP.	IND.	DEP.	IND.	DEP.		IND.	DEP.	IND.	DEP.	IND.	DEP.
1							51						
2							52						
3							53						
4							54						
5							55						
6							56						
7							57						
8							58						
9							59						
10							60						
11							61						
12							62						
13							63						
14							64						
15							65						
16							66						
17							67						
18							68						
19							69						
20							70						
21							71						
22							72						
23							73						
24							74						
25							75						
26							76						
27							77						
28							78						
29							79						
30							80						
31							81						
32							82						
33							83						
34							84						
35							85						
36							86						
37							87						
38							88						
39							89						
40							90						
41							91						
42							92						
43							93						
44							94						
45							95						
46							96						
47							97						
48							98						
49							99						
50							100						
TOTAL IND.	5						TOTAL IND.						
TOTAL DEP.	20						TOTAL DEP.						
TOTAL CLAIMS	25						TOTAL CLAIMS						

* MAY BE USED FOR ADDITIONAL CLAIMS OR AMENDMENTS

FORM PTO-2022 (1-98)
U.S. DEPARTMENT OF COMMERCE
Patent and Trademark Office

BEST AVAILABLE COPY

*U.S. GPO: 1998-443-593/89152



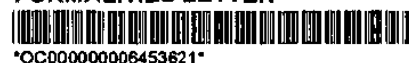
UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

APPLICATION NUMBER	FILING/RECEIPT DATE	FIRST NAMED APPLICANT	ATTORNEY DOCKET NUMBER
09/898,289	07/03/2001	Geoffrey G. Zweig	YOR920010253US1

CONFIRMATION NO. 2216

FORMALITIES LETTER



OC000000006453621

Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560

Date Mailed: 08/21/2001

NOTICE TO FILE MISSING PARTS OF NONPROVISIONAL APPLICATION

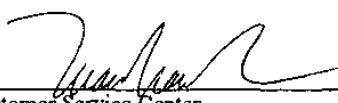
FILED UNDER 37 CFR 1.53(b)

Filing Date Granted

An application number and filing date have been accorded to this application. The item(s) indicated below, however, are missing. Applicant is given **TWO MONTHS** from the date of this Notice within which to file all required items and pay any fees required below to avoid abandonment. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

- The oath or declaration is unsigned.
- To avoid abandonment, a late filing fee or oath or declaration surcharge as set forth in 37 CFR 1.16(e) of \$130 for a non-small entity, must be submitted with the missing items identified in this letter.
- The balance due by applicant is \$ 130.

*A copy of this notice **MUST** be returned with the reply.*


Customer Service Center
Initial Patent Examination Division (703) 308-1202

PART 3 - OFFICE COPY



UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

APPLICATION NUMBER	FILING/RECEIPT DATE	FIRST NAMED APPLICANT	ATTORNEY DOCKET NUMBER
09/898,289	07/03/2001	Geoffrey G. Zweig	YOR920010253US1

CONFIRMATION NO. 2216

FORMALITIES LETTER



OC00000006453621

Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560

Date Mailed: 08/21/2001

NOTICE TO FILE MISSING PARTS OF NONPROVISIONAL APPLICATION

FILED UNDER 37 CFR 1.53(b)

Filing Date Granted

An application number and filing date have been accorded to this application. The item(s) indicated below, however, are missing. Applicant is given **TWO MONTHS** from the date of this Notice within which to file all required items and pay any fees required below to avoid abandonment. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

- The oath or declaration is unsigned.
- To avoid abandonment, a late filing fee or oath or declaration surcharge as set forth in 37 CFR 1.16(e) of \$130 for a non-small entity, must be submitted with the missing items identified in this letter.
- The balance due by applicant is \$ 130.

*A copy of this notice **MUST** be returned with the reply.*

Customer Service Center

Initial Patent Examination Division (703) 308-1202

PART 2 - COPY TO BE RETURNED WITH RESPONSE

09/07/2001 MAHMED1 00000011 500510 09898289

01 FC:105 130.00 CH



09/06/01

#2

Sec 101

Attorney Docket No. YCR920010253US1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): G.G. Zweig et al.
Docket No.: YOR920010253US1
Serial No.: 09/898,289
Filing Date: July 3, 2001
Group: 2182
Examiner: Unassigned

"Express Mail" Label No. EL793517479US
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated below and is addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231.

Date of Deposit: September 5, 2001

Signature: Senise A. Glaser

Title: Information Extraction From Documents
With Regular Expression Matching

TRANSMITTAL OF MISSING PARTS

Assistant Commissioner for Patents
Washington, D.C. 20231
Box Missing Parts

Sir:

In response to the Notice to File Missing Parts dated August 21, 2001, Applicants submit herewith the following documents to complete the above-referenced patent application:

- (1) Assignment with Recordation Form Cover Sheet;
- (2) Declaration and Power of Attorney;
- (3) Transmittal of Formal Drawings; and
- (4) Formal Drawings (1 sheet).

A copy of the Notice is enclosed herewith. Please charge **International Business Machines Corporation Deposit Account No. 50-0510** the amount of \$130, to cover the surcharge under 37 CFR §1.16(e). In the event of non-payment or improper payment of a required fee, the Commissioner is authorized to charge or to credit **International Business Machines Corporation Deposit Account No. 50-0510** as required to correct the error. A duplicate copy of this letter is enclosed.

Respectfully submitted,

William E. Lewis

William E. Lewis
Reg. No. 39,274
Attorney for Applicant(s)
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, New York 11560
(516) 759-2946

Date: September 5, 2001



Attorney Docket No. YOR920010253US1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): G.G. Zweig et al.
Docket No.: YOR920010253US1
Serial No.: 09/898,289
Filing Date: July 3, 2001
Group: 2182
Examiner: Unassigned

"Express Mail" Label No. EL793517479US

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated below and is addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231.

Date of Deposit: September 5, 2001

Signature: Senise A. Glaser

Title: Information Extraction From Documents
With Regular Expression Matching

TRANSMITTAL OF FORMAL DRAWINGS

Assistant Commissioner for Patents
Washington, D.C. 20231

Attention: Official Draftsperson

Sir:

Applicants submit herewith one (1) sheet of formal drawings to complete the above-referenced patent application.

In the event of non-payment or improper payment of a required fee, the Commissioner is authorized to charge or to credit **International Business Machines Corporation Deposit Account No. 50-0510** as required to correct the error.

Respectfully submitted,

William E. Lewis

Date: September 5, 2001

William E. Lewis
Reg. No. 39,274
Attorney for Applicant(s)
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, New York 11560
(516) 759-2946

FIG. 1

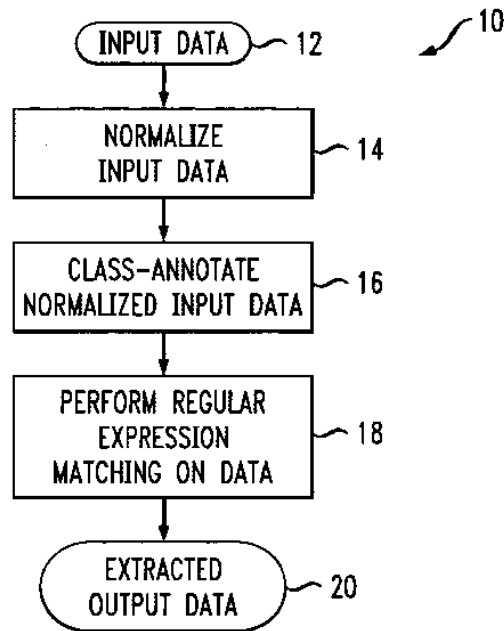
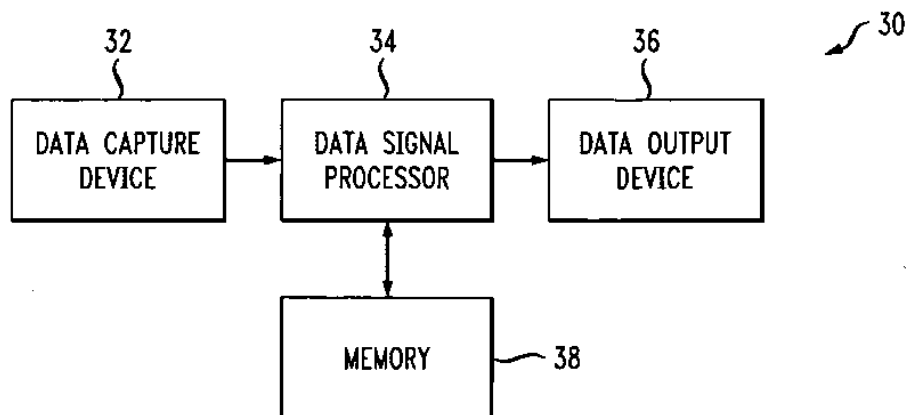


FIG. 2





UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

APPLICATION NUMBER	FILING/RECEIPT DATE	FIRST NAMED APPLICANT	ATTORNEY DOCKET NUMBER
09/898,289	07/03/2001	Geoffrey G. Zweig	YOR920010253US1

CONFIRMATION NO. 2216

FORMALITIES LETTER



OC000000006896865

Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560

Date Mailed: 10/12/2001

NOTICE OF INCOMPLETE REPLY (NONPROVISIONAL)

Filing Date Granted

The U.S. Patent and Trademark Office has received your reply on 09/05/2001 to the Notice to File Missing Parts (Notice) mailed 08/21/2001 and it has been entered into the nonprovisional application. The reply, however, does not include the following items required in the Notice.

The period of reply remains as set forth in the Notice. You may, however, obtain EXTENSIONS OF TIME under the provisions of 37 CFR 1.136 (a) accompanied by the appropriate fee (37 CFR 1.17(a)).

A complete reply must be timely filed to prevent ABANDONMENT of the above-identified application.

The required items noted below **SHOULD** be filed along with any items required above. The filing date of this nonprovisional application will be the date of receipt of the items required above.

- The oath or declaration is unsigned.

*A copy of this notice **MUST** be returned with the reply.*


Customer Service Center
Initial Patent Examination Division (703) 308-1202

PART 3 - OFFICE COPY

DECLARATION

AS A BELOW NAMED INVENTOR, I hereby declare that:

My residence, post office address and citizenship are as stated next to my name.

I believe that I am the original, first and sole (*if only one name is listed below*), or an original, first and joint inventor (*if plural names are listed below*), of the subject matter which is claimed and for which a patent is sought on the invention entitled:
TITLE: INFORMATION EXTRACTION FROM DOCUMENTS WITH REGULAR EXPRESSION MATCHING

the specification of which is attached hereto or indicates an attorney docket no. YOR920010253US1, or:

☒ was filed in the U.S. Patent & Trademark Office on July 3, 2001 and assigned Serial No. 09/898,289,

☐ and (*if applicable*) was amended on _____

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to patentability and to the examination of this application in accordance with Title 37, Code of Federal Regulations §1.56. I hereby claim foreign priority benefits under Title 35, U.S. Code §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate, or §365(a) of any PCT international application which designated at least one country other than the United States, or §119(e) of any United States provisional application(s), listed below and have also identified below any foreign applications for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Priority Claimed:

Yes [] No []

(Application Number)	(Country)	(Day/Month/Year filed)

Yes [] No []

(Application Number)	(Country)	(Day/Month/Year filed)

I hereby claim the benefit under Title 35, U.S. Code §120, of any United States application(s), or §365(c), of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International applications(s) in the manner provided by the first paragraph of Title 35, U.S. Code §112, I acknowledge the duty to disclose information material to patentability as defined in Title 37, Code of Federal Regulations §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

(Application Serial Number)	(Filing Date)	(STATUS: patented, pending, abandoned)

(Application Serial Number)	(Filing Date)	(STATUS: patented, pending, abandoned)

I hereby appoint the following attorneys: MANNY W. SCHECTER, Reg. No. 31,722; LAUREN C. BRUZZONE, Reg. No. 35,082; CHRISTOPHER A. HUGHES, Reg. No. 26,914; JOHN E. HOEL, Reg. No. 26,279; JOSEPH C. REDMOND, JR., Reg. No. 18,753; STANLEY B. GREEN, Reg. No. 24,351; STEPHEN C. KAUFMAN, Reg. No. 29,551; ROBERT M. TREPP, Reg. No. 25,933; LOUIS P. HERZBERG, Reg. No. 41,500; DANIEL P. MORRIS, Reg. No. 32,053; PAUL J. OTTERSTEDT, Reg. No. 37,411; LOUIS J. PERCELLO, Reg. No. 33,206; DOUGLAS W. CAMERON, Reg. No. 31,596; MARIAN UNDERWEISER, Reg. No. 46,134; RICHARD M. LUDWIN, Reg. No. 33,010; MARC A. EHRLICH, Reg. No. 39,966; ROBERT P. TASSINARI, JR., Reg. No. 36,030; DEREK S. JENNINGS, Reg. No. 41,473; GAIL H. ZARICK, Reg. No. 43,303; and TIMOTHY M. FARRELL, Reg. No. 37,321; each of them of INTERNATIONAL BUSINESS MACHINES CORPORATION, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598; to prosecute this application and to transact all business in the U.S. Patent and Trademark Office connected therewith and with any divisional, continuation, continuation-in-part, reissue or re-examination application, with full power of appointment and with full power to substitute an associate attorney or agent, and to receive all patents which may issue thereon, and request that all correspondence be addressed to:

William E. Lewis
 RYAN, MASON & LEWIS, LLP
 90 Forest Avenue
 Locust Valley, NY 11560
 Tel.: (516) 759-2946

I HEREBY DECLARE that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under §1001 of Title 18 U.S. Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

FULL NAME OF FIRST OR SOLE INVENTOR: Geoffrey G. Zweig Citizenship U.S.A.

Inventor's signature: Geoffrey G. Zweig Date: 7/20/01

Residence & Post Office address: 30 Brookside Drive, Apt. 1A
Greenwich, Connecticut 06830

FULL NAME OF SECOND JOINT INVENTOR: Mukund Padmanabhan Citizenship India

Inventor's signature: [Signature] Date: 7/26/01

Residence & Post Office address: 19 Old Mamaroneck Road, #5E
White Plains, New York 10605



UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

APPLICATION NUMBER	FILING/RECEIPT DATE	FIRST NAMED APPLICANT	ATTORNEY DOCKET NUMBER
09/898,289	07/03/2001	Geoffrey G. Zweig	YOR920010253US1

CONFIRMATION NO. 2216

FORMALITIES LETTER

FORMALITIES LETTER
"0C00000006453621"

Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560

Date Mailed: 08/21/2001

NOTICE TO FILE MISSING PARTS OF NONPROVISIONAL APPLICATION

FILED UNDER 37 CFR 1.53(b)

Filing Date Granted

An application number and filing date have been accorded to this application. The item(s) indicated below, however, are missing. Applicant is given **TWO MONTHS** from the date of this Notice within which to file all required items and pay any fees required below to avoid abandonment. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

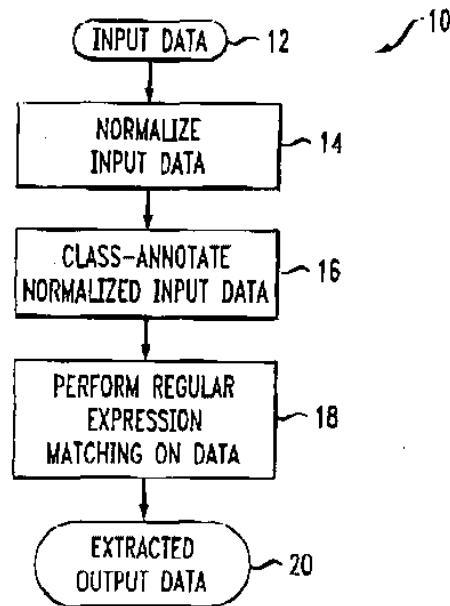
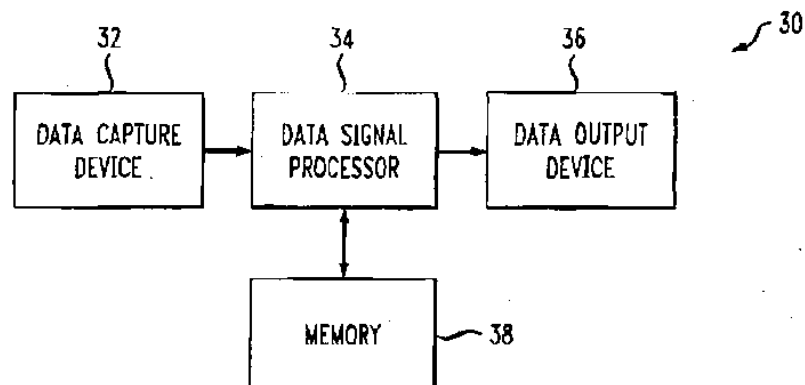
- The oath or declaration is unsigned.
- To avoid abandonment, a late filing fee or oath or declaration surcharge as set forth in 37 CFR 1.16(e) of \$130 for a non-small entity, must be submitted with the missing items identified in this letter.
- The balance due by applicant is \$ 130.

*A copy of this notice **MUST** be returned with the reply.*

Customer Service Center
Initial Patent Examination Division (703) 308-1202

PART 2 - COPY TO BE RETURNED WITH RESPONSE

1/1
ZWEIG et al.
YOR920010253US1 (PJO)

FIG. 1*FIG. 2*



COPIES OF PAPERS
ORIGINALLY FILED

2182⁴⁵
02
01.27.02

Attorney Docket No. YOR920010253US1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): G.G. Zweig et al.
Docket No.: YOR920010253US1
Serial No.: 09/898,289
Filing Date: July 3, 2001
Group: 2182
Examiner: Unassigned

I hereby certify that this paper is being deposited on this date with the
U.S. Postal Service as priority mail addressed to the Assistant
Commissioner for Patents, Washington, D.C. 20231.

Signature:

Date: November 21, 2001

Title: Information Extraction From Documents
With Regular Expression Matching

LETTER

Assistant Commissioner for Patents
Washington, D.C. 20231

RECEIVED

FEB 05 2002

Technology Center 2100

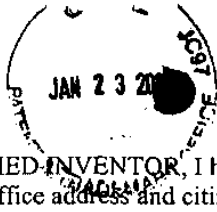
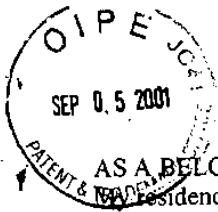
Further to the direction of Pierre Bataille of the OIPE, Group Art Unit 2600, we are enclosing an original executed Declaration for the above-referenced application which had been previously filed with the U.S. Patent and Trademark Office on July 3, 2001, in response to an outstanding Notice to File Missing Parts dated August 21, 2001. The Declaration had erroneously been forwarded to us with a Notice of Recordation of Assignment Document. We would appreciate it if you could match this Declaration with the file, and call the undersigned attorney at (516) 759-2946 to confirm that this has been done.

Please let me know if you require any additional information.

Respectfully submitted,

Date: November 21, 2001

William E. Lewis
Attorney for Applicant(s)
Reg. No. 39,274
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-2946



Attorney Docket No. YOR920010253US1

DECLARATION

AS A ~~BELOW NAMED INVENTOR~~, I hereby declare that:

~~My~~ residence, post office address and citizenship are as stated next to my name.

I believe that I am the original, first and sole (*if only one name is listed below*), or an original, first and joint inventor (*if plural names are listed below*), of the subject matter which is claimed and for which a patent is sought on the invention entitled:

TITLE: INFORMATION EXTRACTION FROM DOCUMENTS WITH REGULAR EXPRESSION MATCHING

the specification of which is attached hereto or indicates an attorney docket no. YOR920010253US1, or:

☒ was filed in the U.S. Patent & Trademark Office on July 3, 2001 and assigned Serial No. 09/898,289,

☐ and (*if applicable*) was amended on _____.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to patentability and to the examination of this application in accordance with Title 37, Code of Federal Regulations §1.56. I hereby claim foreign priority benefits under Title 35, U.S. Code §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate, or §365(a) of any PCT international application which designated at least one country other than the United States, or §119(e) of any United States provisional application(s), listed below and have also identified below any foreign applications for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Priority Claimed:

Yes [] No []

(Application Number) (Country) (Day/Month/Year filed)

Yes [] No []

(Application Number) (Country) (Day/Month/Year filed)

I hereby claim the benefit under Title 35, U.S. Code §120, of any United States application(s), or §365(c), of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International applications(s) in the manner provided by the first paragraph of Title 35, U.S. Code §112, I acknowledge the duty to disclose information material to patentability as defined in Title 37, Code of Federal Regulations §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

(Application Serial Number) (Filing Date) (STATUS: patented, pending, abandoned)

(Application Serial Number) (Filing Date) (STATUS: patented, pending, abandoned)

I hereby appoint the following attorneys: **MANNY W. SCHECTER**, Reg. No. 31,722; **LAUREN C. BRUZZONE**, Reg. No. 35,082; **CHRISTOPHER A. HUGHES**, Reg. No. 26,914; **JOHN E. HOEL**, Reg. No. 26,279; **JOSEPH C. REDMOND, JR.**, Reg. No. 18,753; **STANLEY B. GREEN**, Reg. No. 24,351; **STEPHEN C. KAUFMAN**, Reg. No. 29,551; **ROBERT M. TREPP**, Reg. No. 25,933; **LOUIS P. HERZBERG**, Reg. No. 41,500; **DANIEL P. MORRIS**, Reg. No. 32,053; **PAUL J. OTTERSTEDT**, Reg. No. 37,411; **LOUIS J. PERCELLO**, Reg. No. 33,206; **DOUGLAS W. CAMERON**, Reg. No. 31,596; **MARIAN UNDERWEISER**, Reg. No. 46,134; **RICHARD M. LUDWIN**, Reg. No. 33,010; **MARC A. EHRLICH**, Reg. No. 39,966; **ROBERT P. TASSINARI, JR.**, Reg. No. 36,030; **DEREK S. JENNINGS**, Reg. No. 41,473; **GAIL H. ZARICK**, Reg. No. 43,303; and **TIMOTHY M. FARRELL**, Reg. No. 37,321; each of them of **INTERNATIONAL BUSINESS MACHINES CORPORATION**, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598; to prosecute this application and to transact all business in the U.S. Patent and Trademark Office connected therewith and with any divisional, continuation, continuation-in-part, reissue or re-examination application, with full power of appointment and with full power to substitute an associate attorney or agent, and to receive all patents which may issue thereon, and request that all correspondence be addressed to:

William E. Lewis
RYAN, MASON & LEWIS, LLP
90 Forest Avenue
Locust Valley, NY 11560
Tel.: (516) 759-2946

RECEIVED

FEB 05 2002

Technology Center 2100

I HEREBY DECLARE that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under §1001 of Title 18 U.S. Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

FULL NAME OF FIRST OR SOLE INVENTOR: Geoffrey G. Zweig Citizenship U.S.A.

Inventor's signature: *Geoffrey G. Zweig* Date: 7/20/01
Residence & Post Office address: 30 Brookside Drive, Apt. 1A
Greenwich, Connecticut 06830

FULL NAME OF SECOND JOINT INVENTOR: Mukund Padmanabhan Citizenship India

Inventor's signature: *Mukund Padmanabhan* Date: 7/26/01
Residence & Post Office address: 19 Old Mamaroneck Road, #5E
White Plains, New York 10605



COPY OF PAPERS
ORIGINALLY FILED

Attorney Docket No. YOR920010253US1

2182
ab
5-31-02

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): G.G. Zweig et al.
Docket No.: YOR920010253US1
Serial No.: 09/898,289
Filing Date: July 3, 2001
Group: 2182
Examiner: Unassigned

I hereby certify that this paper is being deposited on this date with the U.S. Postal Service as first class mail addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.

Signature: Senise A. Gleser Date: May 8, 2002

Title: Information Extraction From Documents
With Regular Expression Matching

RECEIVED

INFORMATION DISCLOSURE STATEMENT

MAY 23 2002

Technology Center 2100

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Pursuant to 37 C.F.R. §§1.56, 1.97 and 1.98, Applicants' attorney wishes to bring to the attention of the Patent and Trademark Office the following document listed on the accompanying Form PTO-1449 PTO-1449. A copy of each listed document is enclosed.

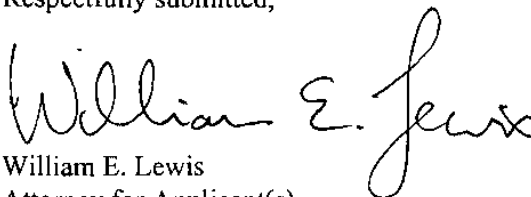
Other Documents

1. Ratnaprkhi, "A Maximum Entropy Part of Speech Tagger," Conference on Empirical Methods in Natural Language Processing, University of Pennsylvania, 1996.
2. Brill, "Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging," Computational Linguistics, December 1995.

It is believed that there is no fee due in conjunction with the filing of this Information Disclosure Statement. In the event of non-payment or improper payment of a required fee, the Commissioner is authorized to charge or to credit **International Business Machines Corporation Deposit Account No. 50-0510** as required to correct the error.

The filing of this Information Disclosure Statement shall not be construed as a representation that a search has been made, or as an admission that the information cited is considered to be material to patentability, or as a representation that no other material information exists.

Respectfully submitted,

A handwritten signature in cursive script that reads "William E. Lewis". The signature is written in dark ink and is positioned above the printed name and address.

Date: May 8, 2002

William E. Lewis
Attorney for Applicant(s)
Reg. No. 39,274
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-2946

L Number	Hits	Search Text	DB	Time stamp
1	276	710/65.ccls.	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2003/07/29 16:07
2	7	(identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2) same document\$2 and 710/65.ccls.	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2003/07/29 16:09
3	0	(extract\$4) with (data or symbol\$2 or expression\$2) with document\$2 and 710/65.ccls.	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2003/07/29 16:10
-	17846	(identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2) same document\$2	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2003/07/29 16:08
-	6279	(extract\$4) same (data or symbol\$2 or word\$2 same expression\$2) same document\$2	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2003/07/28 13:14
-	3242	(extract\$4) with (data or symbol\$2 or word\$2 same expression\$2) with document\$2	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2003/07/28 13:34
-	3330	(extract\$4) with (data or symbol\$2 or expression\$2) with document\$2	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2003/07/29 16:10
-	8	(extract\$4) with (data adj2 symbol\$2) with document\$2	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2003/07/28 13:17
-	3143	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with document\$2) and g06f\$.ipc.	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2003/07/28 13:39
-	13	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with (voice near4 message)) and g06f\$.ipc.	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2003/07/28 13:55
-	8	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with (voice near4 mail)) and g06f\$.ipc.	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2003/07/28 14:11
-	0	20030050782.URPN.	USPAT	2003/07/28 13:57
-	43	((extract\$4) with ((key adj2 word\$2) or information\$3)) with (voice near4 (messag\$3 or mail\$3)) and g06f\$.ipc.	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2003/07/28 14:14

Search History 7/29/03 4:11:53 PM Page 1

C:\APPS\EAST\Workspaces\09898289.wsp



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/898,289	07/03/2001	Geoffrey G. Zweig	YOR920010253US1	2216

7590 08/04/2003
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560

EXAMINER

KNAPP, JUSTIN R

ART UNIT	PAPER NUMBER
----------	--------------

2182

DATE MAILED: 08/04/2003

7

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 09/898,289	Applicant(s) ZWEIG ET AL.	
	Examiner Justin Knapp	Art Unit 2182	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) ☒ Responsive to communication(s) filed on 03 July 2001.

2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.

3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) ☒ Claim(s) 1-25 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5) ☐ Claim(s) _____ is/are allowed.

6) ☒ Claim(s) 1-25 is/are rejected.

7) ☐ Claim(s) _____ is/are objected to.

8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) ☐ The specification is objected to by the Examiner.

10) ☒ The drawing(s) filed on 05 September 2001 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

11) ☐ The proposed drawing correction filed on _____ is: a) ☐ approved b) ☐ disapproved by the Examiner.
If approved, corrected drawings are required in reply to this Office action.

12) ☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

13) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
* See the attached detailed Office action for a list of the certified copies not received.

14) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).
a) ☐ The translation of the foreign language provisional application has been received.

15) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)	4) ☐ Interview Summary (PTO-413) Paper No(s). _____
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)	5) ☐ Notice of Informal Patent Application (PTO-152)
3) ☒ Information Disclosure Statement(s) (PTO-1449) Paper No(s) <u>6</u> .	6) ☐ Other: _____

DETAILED ACTION

Papers Submitted

1. It is hereby acknowledged that the following papers have been received and placed of record in the file: Information Disclosure Statement as received 05/20/02; Declaration as received 01/23/02; Drawings as received 09/05/01.

Claim Rejections - 35 USC § 102

2. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

3. Claims 1, 3-8, 10-12, 14-19, and 21-24 are rejected under 35 U.S.C. 102(e) as being anticipated by Adler, et al (herein referred to as Adler), US Patent Number 6,342,901.

4. Referring to method claim 1, apparatus claim 12, and manufacture claim 23, Adler teaches at least one processor operative to: (i) identify at least one regularly identifiable expression in the input sequence of data symbols; (ii) identify at least a portion of information associated with the at least one regularly identifiable expression; and (iii) extract the portion of information; and memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations.

To further explain, Adler teaches that cue words which are “regularly identifiable expressions” (column 16, lines 10-20 and column 17, lines 40-51) are used to identify portions of information

within a message (such as the person's name or phone number). Pointers are then used to mark those parts of the message. The parts of the message pointed to can then be extracted and stored (see Figure 13, elements 1312 through 1332 and column 15, line 63 through column 20, line 59).

5. Referring to method claim 3 and apparatus claim 14, Adler teaches wherein the at least one processor is further operative to normalize the input sequence of data symbols prior to identifying the regularly identifiable expression. Adler teaches that when a document is scanned, handwritten messages are converted or normalized to a machine-readable format and then stored in a database. They can then be identified. (see column 13, lines 22-36)

6. Referring to method claim 4 and apparatus claim 15, Adler teaches wherein the at least one processor is further operative to identify one or more classes of data symbols in the input sequence prior to identifying the regularly identifiable expression. A variety of classes such as a name class or a phone number class are identifiable and stored in a database.

7. Referring to method claim 5 and apparatus claim 16, Adler teaches wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically precedes a particular portion of information. (see column 20, lines 45-56)

8. Referring to method claim 6 and apparatus claim 17, Adler teaches wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically follows a particular portion of information. (see column 20, lines 45-56)

9. Referring to method claim 7 and apparatus claim 18, Adler teaches wherein the extracted portion of information is used to take a specified action. Adler teaches a number of actions that can be performed once the information is extracted. (see column 17, lines 18-31)

10. Referring to method claim 8 and apparatus claim 19, Adler teaches wherein the extracted portion of information is at least one of visually and audibly presented to the user. Figure 13, elements 1326, 1328, and 1332 show that extracted information is playable for the user to listen to.

11. Referring to method claim 10 and apparatus claim 21, wherein the input sequence of data symbols is representative of at least one of text data, transcribed spoken data, deoxyribonucleic acid sequence data, ribonucleic acid sequence data, amino-acid sequence data, audio sequence data, and video sequence data. As taught herein above, Adler teaches an input sequence of audio sequence data.

12. Referring to method claim 11 and apparatus claim 22, wherein the input sequence of data symbols is representative of a voice mail message. (see column 18, examples 1, 2, and 3)

13. Referring to claim 24, Adler teaches an apparatus comprising:

a data capture device for obtaining the input sequence of data symbols (see figure 2, elements 226, 228, 238);

at least one processor, operatively coupled to the data capture device, operative to: (i) identify at least one regularly identifiable expression in the input sequence of data symbols; (ii) identify at least a portion of information associated with the at least one regularly identifiable expression; and (iii) extract the portion of information (as taught supra);

memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations (see figure 2, element 214); and

a data output device, operatively coupled to the at least one processor, for presenting the extracted portion of information to a user (see figure 2, elements 230 and 216).

Claim Rejections - 35 USC § 103

14. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

15. Claim 2, 9, 13, 20, and 25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Adler.

16. Referring to method claims 2 and 25 and apparatus claim 13, Adler does not explicitly teach wherein the regularly identifiable expression identifying operation comprises comparing the input sequence of data symbols to one or more regularly identifiable expressions, previously stored in the memory, to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions. However, Adler, as taught supra, uses cue words to identify portions of a message. It would have been obvious to one of ordinary skill in the art at the time the invention was made that Adler's cue words would have to be stored in memory ahead of time and used for comparing incoming data for the purpose of identifying whether input data is a cue word or not.

17. Referring to method claim 9 and apparatus claim 20, Adler does not explicitly teach wherein the regularly identifiable expression identifying operation is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language. Adler does not teach what programming languages are used in the system taught. However, it would have been obvious to one of ordinary skill in the art that identifying operation programs

would be able to be written in a variety of programming languages for the purpose of adding flexibility to the system.

Conclusion

18. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Applicant is reminded that in amending in response to a rejection of claims, the patentable novelty must be clearly shown in view of the state of the art disclosed by the references cited and the objections made. Applicant must also show how the amendments avoid such references and objections.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Justin Knapp whose telephone number is (703) 308-6132. The examiner can normally be reached on Mon - Fri 9 am - 5:30 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Jeffrey Gaffin can be reached on (703) 308-3301. The fax phone numbers for the organization where this application or proceeding is assigned are (703) 746-7239 for regular communications and (703) 746-7238 for After Final communications.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (703) 305-3900.

Justin Knapp
Examiner
Art Unit 2182

Application/Control Number: 09/898,289
Art Unit: 2182

Page 7

jk
July 29, 2003



JEFFREY GAFFIN
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100

Notice of References Cited	Application/Control No. 09/898,289	Applicant(s)/Patent Under Reexamination ZWEIG ET AL.	
	Examiner Justin Knapp	Art Unit 2182	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-6,342,901	01-2002	Adler et al.	345/700
	B	US-6,317,762	11-2001	Okawa et al.	715/541
	C	US-6,002,491	12-1999	Li et al.	358/436
	D	US-2002/0059266	05-2002	l'anson e tal.	707/100
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

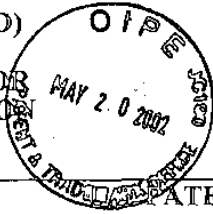
NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

FORM PTO-1449 (MODIFIED)

LIST OF PUBLICATIONS FOR
APPLICANT'S INFORMATION
DISCLOSURE STATEMENT



Applicant(s): G.G. Zweig et al.
Docket No.: YOR920010253US1
Serial No.: 09/898,289
Filing Date: July 3, 2001
Group: 2182

COPY OF PAPER
ORIGINALLY FILED

PATENT DOCUMENTS				
EXAMINER	DOCUMENT NO.	DATE	NAME	CLASS/SUBCLASS
INITIAL				

FOREIGN PATENT DOCUMENTS				
EXAMINER	DOCUMENT NO.	DATE	COUNTRY	TRANSLATION
INITIAL				YES NO

OTHER DOCUMENTS		
EXAMINER	REF. NO.	AUTHOR, TITLE, DATE, PERTINENT PAGES, ETC.
INITIAL		

1. Ratnaprkhi, "A Maximum Entropy Part of Speech Tagger," Conference on Empirical Methods in Natural Language Processing, University of Pennsylvania, 1996.
2. Brill, "Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging," Computational Linguistics, December 1995.

RECEIVED

MAY 23 2002

Technology Center 2100

Examiner

Date Considered

7/29/03

Examiner: Initial if reference considered, whether or not citation is in conformance with MPEP 609; draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

Page 1 of 1



2182

Attorney Docket No. YOR920010253US1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): G.G. Zweig et al.
Docket No.: YOR920010253US1
Serial No.: 09/898,289
Filing Date: July 3, 2001
Group: 2182
Examiner: Justin R. Knapp

I hereby certify that this paper is being deposited on this date with the U.S. Postal Service as first class mail addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Signature: W. Benckman Date: November 4, 2003

Title: Information Extraction from Documents
with Regular Expression Matching

RECEIVED

NOV 10 2003

AMENDMENT TRANSMITTAL

Technology Center 2100

Mail Stop Non-Fee Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Submitted herewith is the following document relating to the above-identified patent application:

(1) Amendment and Response to Office Action.

There is no additional fee due in conjunction with the Amendment. In the event of non-payment or improper payment of a required fee, the Commissioner is authorized to charge or to credit **International Business Machines Corporation Deposit Account No. 50-0510** as required to correct the error.

Respectfully submitted,

William E. Lewis

William E. Lewis
Attorney for Applicant(s)
Reg. No. 39,274
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-2946

Date: November 4, 2003



Attorney Docket No. YOR920010253US1

CRS / Jmida A
SA
11/15/03

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): G.G. Zweig et al.
Docket No.: YOR920010253US1
Serial No.: 09/898,289
Filing Date: July 3, 2001
Group: 2182
Examiner: Justin R. Knapp

I hereby certify that this paper is being deposited on this date with the U.S. Postal Service as first class mail addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Signature: *V. Benavente* Date: November 4, 2003

RECEIVED

NOV 10 2003

Technology Center 2100

Title: Information Extraction from Documents
with Regular Expression Matching

AMENDMENT AND RESPONSE TO OFFICE ACTION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In response to the outstanding Office Action dated August 4, 2003, please amend the above-identified application as follows:

IN THE CLAIMS

A listing of the claims in the present application is as follows:

1. (Currently Amended) A method of automatically processing an input sequence of data symbols, the method comprising the steps of:

identifying at least one regularly identifiable expression in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression;

identifying at least a portion of information associated with the at least one regularly identifiable expression; and

extracting the portion of information.

2. (Original) The method of claim 1, wherein the regularly identifiable expression identifying step comprises comparing the input sequence of data symbols to one or more previously-stored regularly identifiable expressions to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions.

3. (Original) The method of claim 1, further comprising the step of normalizing the input sequence of data symbols prior to identifying the regularly identifiable expression.

4. (Original) The method of claim 1, further comprising the step of identifying one or more classes of data symbols in the input sequence prior to identifying the regularly identifiable expression.

5. (Original) The method of claim 1, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically precedes a particular portion of information.

6. (Original) The method of claim 1, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically follows a particular portion of information.

7. (Original) The method of claim 1, wherein the extracted portion of information is used to take a specified action.

8. (Original) The method of claim 1, wherein the extracted portion of information is at least one of visually and audibly presented to the user.

9. (Original) The method of claim 1, wherein the regularly identifiable expression identifying step is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language.

10. (Original) The method of claim 1, wherein the input sequence of data symbols is representative of at least one of text data, transcribed spoken data, deoxyribonucleic acid sequence data, ribonucleic acid sequence data, amino-acid sequence data, audio sequence data, and video sequence data.

11. (Original) The method of claim 1, wherein the input sequence of data symbols is representative of a voice mail message.

12. (Currently Amended) Apparatus for automatically processing an input sequence of data symbols, the apparatus comprising:

at least one processor operative to: (i) identify at least one regularly identifiable expression in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such

a regularly identifiable expression; (ii) identify at least a portion of information associated with the at least one regularly identifiable expression; and (iii) extract the portion of information; and

memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations.

13. (Original) The apparatus of claim 12, wherein the regularly identifiable expression identifying operation comprises comparing the input sequence of data symbols to one or more regularly identifiable expressions, previously stored in the memory, to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions.

14. (Original) The apparatus of claim 12, wherein the at least one processor is further operative to normalize the input sequence of data symbols prior to identifying the regularly identifiable expression.

15. (Original) The apparatus of claim 12, wherein the at least one processor is further operative to identify one or more classes of data symbols in the input sequence prior to identifying the regularly identifiable expression.

16. (Original) The apparatus of claim 12, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically precedes a particular portion of information.

17. (Original) The apparatus of claim 12, wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically follows a particular portion of information.

18. (Original) The apparatus of claim 12, wherein the extracted portion of information is used to take a specified action.

AM 19. (Original) The apparatus of claim 12, wherein the extracted portion of information is at least one of visually and audibly presented to the user.

20. (Original) The apparatus of claim 12, wherein the regularly identifiable expression identifying operation is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language.

21. (Original) The apparatus of claim 12, wherein the input sequence of data symbols is representative of at least one of text data, transcribed spoken data, deoxyribonucleic acid sequence data, ribonucleic acid sequence data, amino-acid sequence data, audio sequence data, and video sequence data.

22. (Original) The apparatus of claim 12, wherein the input sequence of data symbols is representative of a voice mail message.

23. (Currently Amended) An article of manufacture for use in automatically processing an input sequence of data symbols, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

identifying at least one regularly identifiable expression in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression;

identifying at least a portion of information associated with the at least one regularly identifiable expression; and

extracting the portion of information.

24. (Currently Amended) Apparatus for automatically processing an input sequence of data symbols, the apparatus comprising:

 a data capture device for obtaining the input sequence of data symbols;

at least one processor, operatively coupled to the data capture device, operative to: (i) identify at least one regularly identifiable expression in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression; (ii) identify at least a portion of information associated with the at least one regular expression; and (iii) extract the portion of information;

memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations; and

a data output device, operatively coupled to the at least one processor, for presenting the extracted portion of information to a user.

25. (Currently Amended) A method of automatically processing an input document, the method comprising the steps of:

identifying one or more regular expressions in the input document by comparing the input document to one or more previously-stored regular expressions to determine if there is a match between a portion of the input document and at least one of the previously-stored regular expressions, wherein a regular expression represents a pattern that is matchable in accordance with a programming language that supports such a regular expression; and

identifying at least a portion of information associated with the at least one regular expression for extraction.

REMARKS

The present application was filed on July 3, 2001 with claims 1-25. Claims 1, 12, 23, 24 and 25 are independent. In the outstanding Office Action, the Examiner: (i) rejected claims 1, 3-8, 10-12, 14-19 and 21-24 under 35 U.S.C. §102(e) as being anticipated by U.S. Patent No. 6,342,901 to Adler et al. (hereinafter "Adler"); and (ii) rejected claims 2, 9, 13, 20 and 25 under 35 U.S.C. §103(a) as being unpatentable over Adler.

While Applicants believe that the claims as originally filed are patentably distinguishable over Adler, Applicants have amended independent claims 1, 12, 23, 24 and 25 in a sincere effort to expedite the present application through to issuance. More particularly, Applicants have added the language wherein the at least one regularly identifiable expression ("regular expression" in claim 25) represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression. Support for this amendment may be found throughout the present specification, for example, see page 4, line 23, through page 5, line 6.

An example of such a pattern is provided on page 5, lines 20-23, of the present specification: "[f]or example, a set of personal names such as those found in a phone book may be identified as a 'person-name' class. Subsequently, patterns of the form 'Hi, it's <person-name>+' can be matched, and the words matching the "person-name" tokens can be displayed as the caller."

The Office Action contends that the "cue words" in Adler are the same as the regularly identifiable expressions of the claimed invention. However, this is not accurate. As the claimed invention now expressly recites, a regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression. The "cue words" of Adler are merely discrete words that are part of a limited vocabulary that are recognizable by a voice recognition system (column 16, lines 5-20, of Adler), and thus do not represent patterns that are matchable in accordance with a programming language that supports such patterns, as in the claimed invention. Advantageously, such patterns of the claimed invention may encompass more than a limited vocabulary of cue words associated with a voice recognition system.

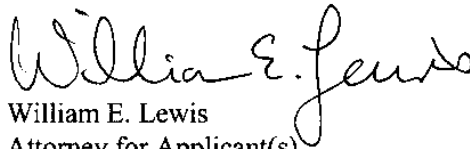
The discrete cue word recognition used by Adler is effectively similar to the conventional word tagging approach discussed in the background section of the present specification, e.g., page 1, line 18, through page 2, line 3. That is, since Adler employs discrete cue words recognizable by a voice recognition system, Adler does not teach or suggest that the cue words can be stored in one of several common programming languages, e.g., as a “flex,” “lex” or “perl” program, as is possible in accordance with the claimed invention. Such programming languages have built-in support for regular expressions. Secondly, since Adler employs discrete cue words recognizable by a voice recognition system, Adler does not teach or suggest pattern matching and information extraction using highly optimized standard programs, as is possible in accordance with the claimed invention. Thirdly, since Adler employs discrete cue words recognizable by a voice recognition system, Adler does not teach or suggest that the cue words can be identified without the expensive and time consuming step of gathering and annotating a “training” database, as is possible in accordance with the claimed invention. It is known that voice recognition systems, such as those used by Adler to recognize cue words, must typically be trained using some appropriate training corpus. Thus, since Adler employs discrete cue words recognizable by a voice recognition system, Adler teaches away from the principles of the claimed invention.

For at least the above reasons, Applicants respectfully assert that independent claims 1, 12, 23, 24 and 25 are patentable over Adler. Further, it is respectfully asserted that dependent claims 2-11 and 13-22 are patentable over Adler not only due to their respective dependence from independent claims 1 and 12, but also because such claims recite patentable subject matter in their own right. By way of example only, claims 9 and 20 recite that the regularly identifiable expression identifying operation is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language. Despite the contention to the contrary in the Office Action, Adler does not teach or suggest this feature since Adler employs a voice recognition system.

Attorney Docket No. YOR920010253US1

In view of the above, Applicants believe that claims 1-25 are in condition for allowance, and respectfully request withdrawal of the §102(e) and §103(a) rejections.

Respectfully submitted,

A handwritten signature in cursive script that reads "William E. Lewis".

Date: November 4, 2003

William E. Lewis
Attorney for Applicant(s)
Reg. No. 39,274
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-2946

L Number	Hits	Search Text	DB	Time stamp
1	2448	705/1.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/25 19:42
2	1433	(identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2) and 705/1.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/25 19:43
3	0	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with (voice near4 message)) and g06f\$.ipc.) and 705/1.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/25 19:43
4	34	705/1.ccls. and ((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with document\$2) and g06f\$.ipc.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/25 19:44
-	17846	(identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2) same document\$2	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 14:33
-	6279	(extract\$4) same (data or symbol\$2 or word\$2 same expression\$2) same document\$2	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/28 13:14
-	3242	(extract\$4) with (data or symbol\$2 or word\$2 same expression\$2) with document\$2	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/28 13:34
-	3330	(extract\$4) with (data or symbol\$2 or expression\$2) with document\$2	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/29 16:10
-	8	(extract\$4) with (data adj2 symbol\$2) with document\$2	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/28 13:17
-	3143	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with document\$2) and g06f\$.ipc.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/25 19:44
-	13	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with (voice near4 message)) and g06f\$.ipc.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/25 19:43
-	8	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with (voice near4 mail)) and g06f\$.ipc.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/28 14:11
-	0	20030050782.URPN.	USPAT	2003/07/28 13:57

Search History 1/25/04 7:44:25 PM Page 1

C:\APPS\EAST\Workspaces\09898289.wsp

-	43	((extract\$4) with ((key adj2 word\$2) or information\$3)) with (voice near4 (messag\$3 or mail\$3)) and g06f\$.ipc.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 14:23
-	276	710/65.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/29 16:07
-	7	(identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2) same document\$2 and 710/65.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/29 16:09
-	0	(extract\$4) with (data or symbol\$2 or expression\$2) with document\$2 and 710/65.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/29 16:10
-	14	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with (voice near4 message)) and g06f\$.ipc.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 14:31
-	230	710/16.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 14:31
-	459691	(identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/25 19:43
-	135	710/16.ccls. and ((identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2))	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 14:35
-	0	(710/16.ccls. and ((identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2))) same voice	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 15:00
-	4	(flex or lex or perl) with (expression\$2 with extract\$3)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 15:07
-	0	(flex or lex or perl) with (phrase with extract\$3)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 15:07
-	1	(flex or lex or perl) with (phrase\$2 with extract\$3)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 15:07



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/898,289	07/03/2001	Geoffrey G. Zweig	YOR920010253US1	2216
7590 01/29/2004				
Ryan, Mason & Lewis, LLP 90 Forest Avenue Locust Valley, NY 11560			EXAMINER KNAPP, JUSTIN R.	
			ART UNIT 2182	PAPER NUMBER

DATE MAILED: 01/29/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 09/898,289	Applicant(s) ZWEIG ET AL.	
	Examiner Justin Knapp	Art Unit 2182	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) ☒ Responsive to communication(s) filed on 06 November 2003.

2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.

3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) ☒ Claim(s) 1-25 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5) ☐ Claim(s) _____ is/are allowed.

6) ☒ Claim(s) 1-10, 12-21 and 23-25 is/are rejected.

7) ☒ Claim(s) 11 and 22 is/are objected to.

8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) ☐ The specification is objected to by the Examiner.

10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. §§ 119 and 120

12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
 a) ☐ All b) ☐ Some * c) ☐ None of:
 1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
 * See the attached detailed Office action for a list of the certified copies not received.

13) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application) since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.
 a) ☐ The translation of the foreign language provisional application has been received.

14) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121 since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892) 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948) 3) ☐ Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____	4) ☐ Interview Summary (PTO-413) Paper No(s). _____ 5) ☐ Notice of Informal Patent Application (PTO-152) 6) ☐ Other: _____
---	---

DETAILED ACTION

Claim Rejections - 35 USC § 102

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

2. Claims 1-10, 12-21, and 23-24 are rejected under 35 U.S.C. 102(e) as being anticipated by Klein, et al (herein referred to as Klein), US Patent Application Pub No. US 2002/0194026.

3. Referring to method claim 1, apparatus claim 12, and manufacture claim 23, Klein teaches at least one processor operative to: (i) identify at least one regularly identifiable expression in the input sequence of data symbols, wherein the at least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression; (ii) identify at least a portion of information associated with the at least one regularly identifiable expression; and (iii) extract the portion of information; and memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations.

To further explain, Klein teaches the process of extracting and storing data in one embodiment of the invention in figures 4 through 9 and section [0047]. In a sequence of data symbols that makes up an inputted document, several regularly identifiable expression patterns are identified.

A PERL language regular expression parser is used along with a rule set that includes a library of

regular expressions that define how information is delimited. A portion of the patient information is identified, extracted, and stored.

4. Referring to method claim 2 and apparatus claim 13, Klein teaches wherein the regularly identifiable expression identifying operation comprises comparing the input sequence of data symbols to one or more regularly identifiable expressions, previously stored in the memory, to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions. As seen in section [0047] a rule set includes a library of regular expressions to use to determine if there is a match between a portion of the input sequence and at least one of the previously-stored regularly identifiable expressions.

5. Referring to method claim 3 and apparatus claim 14, Klein teaches wherein the at least one processor is further operative to normalize the input sequence of data symbols prior to identifying the regularly identifiable expression. Klein teaches that when a document is entered, it is normalized by removing all the formatting in the document.

6. Referring to method claim 4 and apparatus claim 15, Klein teaches wherein the at least one processor is further operative to identify one or more classes of data symbols in the input sequence prior to identifying the regularly identifiable expression. Klein teaches a library of rule sets that is itself a class of what regular expressions to identify.

7. Referring to method claim 5 and apparatus claim 16, Klein teaches wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically precedes a particular portion of information. See figures 4-9.

8. Referring to method claim 6 and apparatus claim 17, Klein teaches wherein the at least one regularly identifiable expression comprises a characteristic phrase that typically follows a particular portion of information. See figures 4-9.

9. Referring to method claim 7 and apparatus claim 18, Klein teaches wherein the extracted portion of information is used to take a specified action. Once the information is extracted, it is analyzed using a Care Guidelines Database (see section [0048])

10. Referring to method claim 8 and apparatus claim 19, Klein teaches wherein the extracted portion of information is at least one of visually and audibly presented to the user. Klein teaches a portion of information is visual presented to the user in figures 4-9.

11. Referring to method claim 9 and apparatus claim 20, Klein teaches wherein the regularly identifiable expression identifying operation is performed in accordance with one or more programs written in one of the flex, lex, and perl programming language. As taught herein above, Klein uses the perl programming language.

12. Referring to method claim 10 and apparatus claim 21, wherein the input sequence of data symbols is representative of at least one of text data, transcribed spoken data, deoxyribonucleic acid sequence data, ribonucleic acid sequence data, amino-acid sequence data, audio sequence data, and video sequence data. As taught herein above, Klein teaches an input sequence of text data.

13. Referring to claim 24, Klein teaches an apparatus comprising:
a data capture device for obtaining the input sequence of data symbols (see figure 1);
at least one processor, operatively coupled to the data capture device, operative to: (i) identify at least one regularly identifiable expression in the input sequence of data symbols wherein the at

least one regularly identifiable expression represents a pattern that is matchable in accordance with a programming language that supports such a regularly identifiable expression; (ii) identify at least a portion of information associated with the at least one regularly identifiable expression; and (iii) extract the portion of information (as taught supra);
memory, operatively coupled to the at least one processor, for storing at least a portion of results associated with the identifying and extracting operations (as taught supra); and
a data output device, operatively coupled to the at least one processor, for presenting the extracted portion of information to a user (see figure 9).

14. Referring to claim 25, Klein teaches a method comprising:
identifying one or more regular expressions in the input document by comparing the input document to one or more previously-stored regular expressions to determine if there is a match between a portion of the input document and at least one of the previously-stored regular expressions, wherein a regular expression represents a pattern that is matchable in accordance with a programming language that supports such a regular expression; and
identifying at least a portion of information associated with the at least one regular expression for extraction (the limitations are taught in the rejections of claims 1 and 2).

Allowable Subject Matter

15. Method claim 11 and apparatus claim 22 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Response to Arguments

16. Applicant's arguments with respect to claims 1-25 on pages 7-8 in Amendment A received 11/06/03 have been considered but are moot in view of the new ground(s) of rejection *supra*.

Conclusion

17. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE-MONTH** shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than **SIX MONTHS** from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Justin Knapp whose telephone number is (703) 308-6132. The examiner can normally be reached on Mon - Fri 9 am - 5:30 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Jeffrey Gaffin can be reached on (703) 308-3301. The fax phone number for the organization where this application or proceeding is assigned is (703) 746-7239.

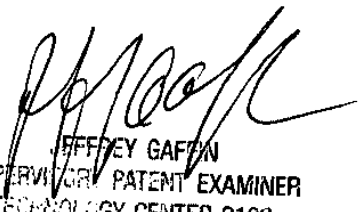
Application/Control Number: 09/898,289
Art Unit: 2182

Page 7

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (703) 305-3900.

Justin Knapp
Examiner
Art Unit 2182

January 25, 2004


JEFFREY GAFFIN
SUPERVISOR, PATENT EXAMINER
TECHNOLOGY CENTER 2100

~~JEFFREY GAFFIN
SUPERVISOR, PATENT EXAMINER
TECHNOLOGY CENTER 2100~~

Notice of References Cited	Application/Control No. 09/898,289	Applicant(s)/Patent Under Reexamination ZWEIG ET AL. 05	
	Examiner Justin Knapp	Art Unit 2182	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-2002/0194026	12-2002	Klein et al.	705/2
	B	US-2002/0010709	01-2002	Culbert et al.	707/500
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.



Attorney Docket No. YOR920010253US1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): G.G. Zweig et al.
Docket No.: YOR920010253US1
Serial No.: 09/898,289
Filing Date: July 3, 2001
Group: 2182
Examiner: Justin R. Knapp

I hereby certify that this paper is being deposited on this date with the U.S. Postal Service as first class mail addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Signature: Lisa L. Chulpis Date: April 29, 2004

Title: Information Extraction from Documents
with Regular Expression Matching

RECEIVED

MAY 05 2004

Technology Center 2100

RESPONSE TO FINAL OFFICE ACTION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In response to the outstanding final Office Action dated January 29, 2004, please consider the following remarks:

REMARKS

The present application was filed on July 3, 2001 with claims 1-25. Claims 1, 12, and 23-25 are independent. In the outstanding final Office Action, the Examiner: (i) rejected claims 1-10, 12-21, 23-25 under 35 U.S.C. §102(e) as being anticipated by U.S. Patent Application Publication No. US 2002/0194026 to Klein et al. (hereinafter "Klein").

Applicants acknowledge the indication of allowable subject matter in claims 11 and 22.

With regard to the rejection of claims 1-10, 12-21 and 23-25 under 35 U.S.C. §102(e) as being anticipated by Klein, Applicants submit herewith a declaration under 37 C.F.R. §1.131. The declaration is signed by the inventors named on the present application. The declaration and the exhibit attached thereto evidence the conception of an invention falling within independent claims 1, 12, 23-25 and at least one or more dependent claims, at least as early as March 14, 2001, and thus prior to the June 13, 2001 effective date of Klein. The declaration further evidences due diligence from March 14, 2001 until the filing date of July 3, 2001.

In view of the above, Applicants believe that claims 1-25 are in condition for allowance, and respectfully request withdrawal of the §102(e) rejection.

Respectfully submitted,



Robert W. Griffith
Attorney for Applicant(s)
Reg. No. 48,956
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-4547

Date: April 29, 2004



YOR920010253US1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): G.G. Zweig et al.
Case: YOR920010253US1
Serial No.: 09/898,289
Filing Date: July 3, 2001
Group: 2182
Examiner: Justin R. Knapp

RECEIVED
MAY 05 2004
Technology Center 2100

Title: Information Extraction from Documents
with Regular Expression Matching

DECLARATION OF PRIOR INVENTION UNDER 37 C.F.R. §1.131

We, the undersigned, hereby declare and state as follows:

1. We are the named inventors on the above-referenced U.S. patent application.
2. The invention that is the subject of the present application was conceived at least as early as March 14, 2001. On or about this date, an IBM disclosure document for an invention entitled "Information Extraction from Documents with Regular Expression Matching" was sent to the inventors' attorneys at the law firm of Ryan, Mason & Lewis, LLP, for preparation of a related patent application. The accompanying letter, dated March 14, 2001, from IBM in-house counsel Paul J. Otterstedt and the IBM disclosure document are attached hereto as Exhibit 1.
3. The IBM disclosure document was written by inventors Geoffrey G. Zweig and Mukund Padmanabhan.

4. Due diligence was performed in the preparation of a patent application from the date the letter and IBM disclosure document were received until the application was filed on July 3, 2001.

5. All statements made herein of our own knowledge are true, and all statements made on information and belief are believed to be true.

6. We understand that willful false statements and the like are punishable by fine or imprisonment, or both, under 18 U.S.C. §1001, and may jeopardize the validity of the application or any patent issuing thereon.

Date: 4/29/04

G. Zweig
Geoffrey G. Zweig

Date: _____

Mukund Padmanabhan



Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

March 14, 2001

William Lewis, Esq.
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, New York 11560

RECEIVED
WITH THANKS
RYAN, MASON & LEWIS, LLP
Mar 14 3:17:01

Re: IBM Docket Number: YOR920010253
Title: "INFORMATION EXTRACTION FROM DOCUMENTS WITH
REGULAR EXPRESSION MATCHING"

Dear Bill:

Please prepare a U.S. Patent Application for filing in the U.S. Patent and Trademark Office for the above referenced docket. This application is to be prepared in accordance with the IBM Outside Counsel Instructions in EPC format.

It is desired that this office be informed at all points involving scope of coverage and financial decisions. We wish that you handle the preparation and filing of all formal papers. Please note that the following paragraph should be used for the Power of Attorney.

POWER OF ATTORNEY: As a named inventor I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith (list name and registration number).

Manny W. Schechter (Reg. 31,722), Lauren Bruzzone (Reg. 35,082), Christopher A. Hughes (Reg. 26,914), John E. Hoel (Reg. 26,279), Joseph C. Redmond, Jr. (Reg. 18,753), Paul J. Otterstedt (Reg. 37,411), Douglas W. Cameron (Reg. 31,596), Stephen C. Kaufman (Reg. 29,551), Daniel P. Morris (Reg. 32,053), Louis J. Percello (Reg. 33,206), Marlan Underweiser (Reg. 46,134), Robert M. Trepp (Reg. 25,933), Louis P. Herzberg (Reg. 41,500) Richard M. Ludwin (Reg. 33,010), Marc A. Ehrlich (Reg. 39,966), Robert P. Tassinari, Jr. (Reg. 36, 030), Derek S. Jennings (Reg. 41, 473), Gail Zarick (Reg. 43,303) and Timothy M. Farrell (Reg. 37,321).

Send Correspondence to: Outside Counsel
Direct Telephone Calls to: (name and telephone number) Outside Counsel

An Associate Power of Attorney form should be prepared and forwarded to this office for signature.

Please be advised that an additional step in our procedure is required when filing ALL original IBM Yorktown Patent Applications in the USPTO. The additional step is that a "Taiwan Oath & Assignment" form must be completed. The form must have all the required information completely filled in and must be signed and dated by all named inventors in the subject patent application. The second page of the two page form has a section entitled, "Note 1", which contains an alphabetized code reference depicting what information must be entered into each corresponding letter field (example: (a), (b), etc.). When the form is complete, and all inventor(s) signatures and dates have been obtained, please forward the original Taiwan Oath & Assignment back to my office.

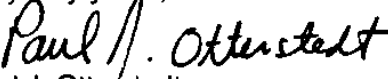
Please have your illustrator prepare formal drawings for the applications. Docket number on drawings must be consistent with docket number on application. In the event these applications are filed with informal drawings, please provide us with formal drawings within six weeks after the filing date. Should you have any questions regarding the preparation of formal drawings, please contact me at (914) 945-3158.

Inventors:

Geoffrey G. Zweig	1-914-945-1204	Yorktown: 23-123
Mukund Padmanabhan	1-914-945-2929	Yorktown: 23-106D

Geoffrey G. Zweig is the Inventor Contact.

Very truly yours,


Paul J. Otterstedt
Attorney - IP Law Dept.

PJO:jd

Enclosures

cc: Barbara Rasa
Geoffrey G. Zweig
Mukund Padmanabhan

Information Extraction from Documents with Regular Expression Matching
Geoffrey Zweig and Mukund Padmanabhan
YOR8-2000-0731

Background of the Invention

1. Field of the Invention

The present invention relates to automatic information extraction from written and spoken documents.

2. Background of the Invention

There are many situations in which it is desirable to extract key pieces of information from documents. For example, in transcribed voicemail messages, the name of the caller and any return numbers that were left are crucial for summarizing the call. Or when resumes are submitted to a company along with a cover letter, it is desirable to extract the job objective and salary requirements of the applicant, in order to determine if a suitable match exists. This invention comprises a simple, efficient, and effective way of extracting such information.

In the past, several techniques have been developed to solve the somewhat simpler problem of "named entity extraction." In this task, the goal is to identify all occurrences of certain classes of words in a document. For example, all the person-names, city-names, dates, and times might be identified. One way of identifying such entities is to train a statistical classification system to tag each word in the document as either a "person-name," "city-name," "date," "time," or "other" word. Examples of this sort of approach can be found in, e.g. Ratnaparkhi ("A Maximum Entropy Part of Speech Tagger," In Eric Brill and Kenneth Church, editors, *Conference on Empirical Methods in Natural Language Processing, University of Pennsylvania*, 1996) and Brill ("Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging," in *Computational Linguistics*, Dec. 1995). While the problem of extracting key pieces of information can also be viewed as a tagging problem - in which each word is tagged as either "key information" or "irrelevant," the present invention takes a more direct approach and exploits the readily-identifiable structure of language to explicitly identify the portions of text that are important.

In common linguistic usage, people usually convey information in highly stereotyped ways. So, for example, in a phone call, the caller is likely to identify himself in one of just a few ways, e.g. "Hi <recipient-name> it's <caller-name>" or "<recipient-name> <caller-name> here." Or in a cover letter, a job applicant is likely to express his interest with a phrase like "I am looking for a job in <field>."

The basic idea of this invention is to enumerate all the stereotypical phrases that people use to convey important information, and where exactly in these phrases the information is to be found. This approach has several advantages over the previously mentioned tagging approach. First, the phrases are readily expressible as regular expressions, and this allows them to be stored in one of several common programming languages, e.g. As a "flex" or "perl" program. Both these languages have built-in support for regular expressions. Secondly, the use of regular-expression technology enables extremely fast pattern matching and information extraction with highly optimized standard programs. Thirdly, the phrases can be identified without the expensive and time consuming step of gathering and annotating a large "training" database. Finally, if desired, the technique can be combined with statistical based procedures, e.g. By using

the phrases as features in a statistical system, or by using a statistical procedure to automatically collect a large pool of candidate phrases which can then be selected for inclusion by a human.

Summary of the Invention

This invention consists of using regular expressions to identify the information-bearing portions of a document, and the explicit identity of the information found therein. Regular expressions (Aho Hopcroft and Ullman) consist of symbolic strings, typically word sequences, that can be constructed by the basic operations of "and" and "or" acting on a atomic alphabet of symbols. So, for example, the pattern "*The dog*" will match only when the symbol the "the" is followed by the symbol "dog." This represents the operation of concatenation or "and." The pattern "*The (dog | cat)*" illustrates the "or" operation: it will match either the string "The dog" or "The cat." The operation of "classing" is also possible, so, for example, the "pet" class may be defined as either "dog" or "cat."

<animal> ::= (dog | cat)

And then subsequent regular expressions can make use of this class, for example, the pattern: "*The <pet>.*" A wide variety of extensions to regular expressions exist, for example, restrictions on where in a document a pattern is found, or restrictions on the length of repeated sequences in a regular expression, and in this patent, the term "regular expression" is taken to mean any sort of pattern that is matchable in the *flex* and *perl* programming languages. Similarly, the term "matching" refers to the processes by which these programs match their input.

In this invention, when an information-bearing regular expression is matched, an action is taken to display the information in an appropriate way. There are two main types of regular expressions that are useful: trigger phrases and trigger suffixes. A trigger prefix is a characteristic phrase that typically precedes a piece of information, for example in a voicemail message, "give me a call back at" is a trigger-phrase that precedes a phone number. A trigger suffix is a phrase that typically follows a key piece of information, even when a trigger prefix is not present. For example the phrase "talk to you later bye" is often a post-facto signal that a phone number has been provided, when the words immediately preceding it are a sequence of numbers. In this invention, trigger prefixes and suffixes may be used either separately, or combined together into larger composite patterns. The operation of word classing is also inherent in this invention. For example, a set of personal names such as those found in a phone book may be identified as a "person-name" class. Subsequently, patterns of the form "*Hi it's <person-name> +*" can be matched, and the words matching the "person-name" tokens can be displayed as the caller. It will be understood by someone skilled in the art that there are numerous ways of combining trigger prefixes and suffixes, word classes, and other pattern-matching operations such as the use of multiple states each with its own set of matchable patterns (common in *flex*) in order to identify and extract information.

In addition to being used as a stand-alone method for information extraction, regular expressions can be in combination with statistical techniques. For example, the existence of a pattern such as "*hi <person-name> it's <person-name>*" can be used as a feature in a maximum-entropy word-tagging system.

Brief Description of the Drawings

Figure 1 illustrates the sequence of operations performed in a preferred embodiment of this invention. Figure 2 presents a *flex* program suitable for information extraction from voicemail messages.

Detailed Description of the Preferred Embodiment

Figure 1 illustrates the information processing performed in a preferred embodiment of this invention. In an initial step, the input is normalized. This consists of standardizing any characteristics of the input that are irrelevant to, or a hindrance to, information extraction. An example of normalization, for text input, is to change all the letters to lower-case and to replace multiple blank spaces between words by a single blank space. This makes subsequent pattern matching easier. In the second step, one or more word (or in general symbol) lists are used to identify and mark occurrences of words that belong to certain classes. For example, all proper names might be marked by prefixing them with an "!" Mark.

In the next step, the normalized and class-annotated text is read by a regular-expression matching program, and the information-bearing portions of the text are analyzed. In a preferred embodiment,

The regular expressions are specified in and matched by a specialized regular expression matching program such as *flex*. In this process, it is understood that all the mechanisms available in such a program may be used. These include the use of special start states, backtracking, context-sensitive matching, and embedded special-purpose "C" code.

Figure 2 presents an illustratory *flex* program for extracting caller names and numbers from voicemail messages. This program is meant to be purely illustratory, and one skilled in the art will be able to create alternative embodiments. Similarly, the same basic regular-expression matching ideas can be used to identify information in other sorts of documents.

It will also be understood that this invention applies to many input forms besides typed text. For example, it applies to information extraction from any data stream that can be symbolically represented, for example: DNA sequences, RNA sequences, amino-acid sequences, and audio and video sequences that are pre-processed into a discrete symbolic form.

What is Claimed is:

1. A process by which a set of regular expressions are used to extract information from documents. In this context, "document" refers to any symbolic sequence and encompasses, for example, written text and DNA sequences.
2. A process based on 1) in which specified actions are taken when regular expressions are matched.
3. A process based on 2) in which the regular expressions include trigger prefixes or trigger suffixes.
4. A process based on 1) in which word classes are used.

5. A process based on 1) in which the input is normalized before the regular expressions are matched.
6. A process based on 1) in which information is extracted from voicemail messages.
7. A process based on 1) in which a flex, lex, perl, or other standardly available program is used to match the regular expressions.

Figure 1: Information Extraction Steps

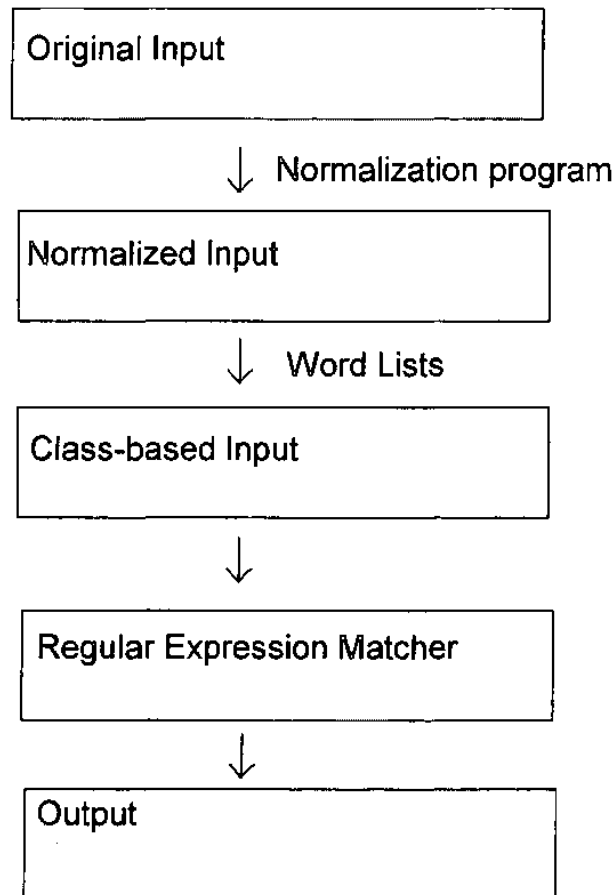


Figure 2: A Sample *flex* Program for Normalizing Text Input

```
%{  
  
#include <iostream.h>  
#include <ctype.h>  
  
%}  
  
%option noyywrap  
%option never-interactive  
  
ws [ \t]+  
  
%%  
  
- {cout << " ";}  
[!\\.@,"()?] {;}  
\\<[^>]*\\>{ws}* {;}  
{ws} {cout << " ";}  
. {cout << char(tolower(*yytext));}  
  
%%  
  
main()  
{  
    yylex();  
}
```

Figure 3: A C++ Program for Identifying Names in a Text Stream

```
#include <iostream.h>
#include <fstream.h>
#include <stdlib.h>
#include <set>
#include <string>

main()
{
    set<string> name;
    ifstream namesin("../data/person_names"); // a text file listing names
    if (!namesin)
    {
        cout << "Unable to find ../data/person_names\n";
        exit(1);
    }

    while (!namesin.eof())
    {
        string s;
        namesin >> s >> ws;
        name.insert(s);
    }
    namesin.close();

    while (!cin.eof())
    {
        string s;
        cin >> s >> ws;
        if (name.count(s))
            cout << "!" << s << " ";
        else
            cout << s << " ";
    }
}
```

Figure 4: A flex Program for Identifying Caller-Names and Phone-Numbers in Voicemail Messages

```
%{

#include <iostream.h>
#include <string.h>

const char *number_type;
const char *tl = "teline ";
const char *extension = "extension ";
const char *pager = "pager ";
const char *fax = "fax number";
const char *none = "";

char number_buff[10000];

}%

%option noyywrap
%option never-interactive

/* many patterns like IN-GENERIC_NAME-AT-DIGIT+ unaccounted for */

spelled_digit
((zero|oh|one|two|three|four|five|six|seven|eight|nine|hundred|thousand|ten|eleven|twelve|thirteen|fourteen|fifteen|sixteen|seventeen|eighteen|nineteen|twenty|thirty|forty|fifty|sixty|seventy|eighty|ninety|hundred)[ ]?)
numeric_digit ((1|2|3|4|5|6|7|8|9|10)[ ]?)
digit ([ ]?{spelled_digit}|[ ]?{numeric_digit})
phone_number ((country[ ]code[ ]{digit}{2,})?(toll[ ]free)?(area[ ]area[ ]code[ ])?{digit}{2,}((or[ ]that's[ ]sorry[ ]){digit}{2,})?(option[ ]{digit}+)?(extension[ ]{digit}{2,})?)
tie_line ([ ]?(tie[ ]line|teline)([ ]number)?)
name ![a-z']+

%x PHONE_NUMBER_STATE
%x NAME_STATE
%x EXTENDED_NAME_STATE
%x TERMINATION_STATE

%%

(1|2|3|4|5|6|7|8|9|10)+": " {cout << yytext << endl; /* get header */ }

{digit}({digit}|"extension"){3,} {strcpy(number_buff, yytext); BEGIN
TERMINATION_STATE;}
```

```

<TERMINATION_STATE>"thank you " |
<TERMINATION_STATE>"thanks " |
<TERMINATION_STATE>"take care " |
<TERMINATION_STATE>"bye " |
<TERMINATION_STATE>"talk to you " |
<TERMINATION_STATE>"okay bye " |
<TERMINATION_STATE>"have a "(good|great)" day " {cout << "Number: " << number_buff
<< endl; BEGIN INITIAL;}
<TERMINATION_STATE><<EOF>> {cout << "Number: " << number_buff << endl; BEGIN
INITIAL;}
<TERMINATION_STATE>. {unput(*yytext); BEGIN INITIAL;}

"on extension " |
"at extension " |
"extension " |
"extension is " {number_type = extension; BEGIN PHONE_NUMBER_STATE;}

"my "{tie_line}" is " |
"my "{tie_line}" here is " |
{tie_line}" is " |
{tie_line}" " {number_type = tl; BEGIN PHONE_NUMBER_STATE;}

"again is " |
"again " |
"back " |
"call me " |
"call me at " |
"call me back at " |
"call me "[a-zA-Z]+" at " |
"call me back on " |
"call me on " |
"call back at " |
"call back on " |
"give me a buzz " |
"give me a buzz at " |
"give me a holler " |
"give me a holler at " |
"give me a ring " |
"give me a call " |
"give me a ring at " |
"give me a call at " |
"give me a call at home " |
"give me a call at work " |
"give me a call at your convenience " |
"give me a call back at " |

```

"give me a call back on " |
 "give me a call back with that information at " |
 "give me a call if you get a chance " |
 "give me a call on " |
 "give me a call please " |
 "give me a call please at " |
 "give me a call when you get a chance " |
 "give me a call when you can " |
 "give me a call when you can please " |
 "give us a call " |
 "give us a call at " |
 "here " |
 "i can be reached at " |
 "i am at " |
 "i'm at " |
 "i'll be at " |
 "i'm on " |
 "leave me a message " |
 "leave me a message at " |
 "my number is " |
 "my number here is " |
 "my number here in "(!?[a-z]+[])"is " |
 "my number's " |
 "my phone number's " |
 "number again is " |
 "number at home is " |
 "number at work is " |
 "number here is " |
 "number is " |
 "number's " |
 "outside line " |
 "outside " |
 "please call " |
 "phone " |
 "please return our call at " |
 "reach me at " |
 "returning your call " |
 "talk to you later " |
 "try me at " |
 "try me at this number " |
 "you could reach us at " |
 "when you have a chance " |
 "when you get a chance " {number_type=none; BEGIN PHONE_NUMBER_STATE;}

"fax number is " |
 "my fax number's" {number_type=fax; BEGIN PHONE_NUMBER_STATE;}

```

"my pager " |
"my pager is " |
"give me a page " |
"give me a page at " |
"page me at " |
"page me " |
"page number "      {number_type=pager; BEGIN PHONE_NUMBER_STATE;}

<PHONE_NUMBER_STATE>{phone_number} {cout << "Number: " << number_type <<
yytext << endl; BEGIN INITIAL;}
<PHONE_NUMBER_STATE>. {unput(*yytext); BEGIN INITIAL;}

"it's "{phone_number} |
"on "{phone_number} {char *c=strchr(yytext, ' '); c++; cout << "Number: " << c << endl;}

"it's me " |
"it's just me "      {cout << "Name: me" << endl;}
"this is your wife " {cout << "Name: wife" << endl;}
"this is receiving " {cout << "Name: receiving" << endl;}
"this is the mail room " {cout << "Name: the mail room" << endl;}

^({name}[ ]){2,3} {cout << "Name: " << (strchr(yytext, ' ')+1) << endl;}
("hi "|"hey "|"hello"){name}" "{name}" "{(name}" " ")? {char *c=strchr(yytext, ' ');
c=strchr(c+1, ' '); cout << "Name: " << (c+1) << endl;}

"good morning "{name}" " |
"yeah "{name}" " |
({name}[ ])? "this is " |
{name}" it's " |
{name}" hi " |
"hi i'm " |
"my name is " |
"my name's " |
"the name is " |
"the name's " |
"hey "({name}[ ])? "this "("is ")? |
"hi "({name}[ ])? "this "("is ")? |
({name}[ ])? "hi it's " |
({name}[ ])? "hey it's " |
({name}[ ])? "how you doing it's " |
{name}" this "      {BEGIN NAME_STATE;}

<NAME_STATE>({name}[ ])+ {cout << "Name: " << yytext << endl; number_type=none;
BEGIN PHONE_NUMBER_STATE;}

```

```
<NAME_STATE>("me "|"mom "|"your sis"|"your sister"|"your wife"|"your brother"|"your
father"|"your husband") {cout << "Name: " << yytext << endl; BEGIN INITIAL;}
```

```
<NAME_STATE>({name}[ ])+"from " |
<NAME_STATE>({name}[ ])+"calling from " |
<NAME_STATE>({name}[ ])+"i'm calling from " |
<NAME_STATE>({name}[ ])+"with " |
<NAME_STATE>({name}[ ])+"i'm with " |
<NAME_STATE>({name}[ ])+"at " |
<NAME_STATE>({name}[ ])+"over at " |
<NAME_STATE>({name}[ ])+"down at " |
<NAME_STATE>({name}[ ])+"over in " |
<NAME_STATE>({name}[ ])+"up in " |
<NAME_STATE>({name}[ ])+"down in " |
<NAME_STATE>({name}[ ])+"up on " |
<NAME_STATE>({name}[ ])+"down on " |
<NAME_STATE>({name}[ ])+"in " {cout << "Name: " << yytext; BEGIN
EXTENDED_NAME_STATE;}
<NAME_STATE>. {unput(*yytext); BEGIN INITIAL;}
```

```
<EXTENDED_NAME_STATE>" i'm " |
<EXTENDED_NAME_STATE>" i " |
<EXTENDED_NAME_STATE>" i've " |
<EXTENDED_NAME_STATE>" i'd " |
<EXTENDED_NAME_STATE>" my " |
<EXTENDED_NAME_STATE>" it's " |
<EXTENDED_NAME_STATE>" it is " |
<EXTENDED_NAME_STATE>" we " |
<EXTENDED_NAME_STATE>" when " |
<EXTENDED_NAME_STATE>" you " |
<EXTENDED_NAME_STATE>" calling " |
<EXTENDED_NAME_STATE>" now " |
<EXTENDED_NAME_STATE>" a " |
<EXTENDED_NAME_STATE>" if " |
<EXTENDED_NAME_STATE>" give " |
<EXTENDED_NAME_STATE>" their " |
<EXTENDED_NAME_STATE>" and " |
<EXTENDED_NAME_STATE>" notifying " |
<EXTENDED_NAME_STATE>" returning " |
<EXTENDED_NAME_STATE>" sorry " |
<EXTENDED_NAME_STATE>" can " |
<EXTENDED_NAME_STATE>" got " |
<EXTENDED_NAME_STATE>" earlier " |
<EXTENDED_NAME_STATE>" your " |
<EXTENDED_NAME_STATE>" just " |
<EXTENDED_NAME_STATE>" that " |
```

```

<EXTENDED_NAME_STATE>" could " |
<EXTENDED_NAME_STATE>" would " |
<EXTENDED_NAME_STATE>" before " |
<EXTENDED_NAME_STATE>" thanks " |
<EXTENDED_NAME_STATE>" thank " |
<EXTENDED_NAME_STATE>" wanted " |
<EXTENDED_NAME_STATE>" following " |
<EXTENDED_NAME_STATE>" regarding " |
<EXTENDED_NAME_STATE>" may " |
<EXTENDED_NAME_STATE>" let " |
<EXTENDED_NAME_STATE>" he " |
<EXTENDED_NAME_STATE>" about " |
<EXTENDED_NAME_STATE>" again " |
<EXTENDED_NAME_STATE>" wanting " |
<EXTENDED_NAME_STATE>" appreciate " |
<EXTENDED_NAME_STATE>" was " |
<EXTENDED_NAME_STATE>" we're " |
<EXTENDED_NAME_STATE>" or " |
<EXTENDED_NAME_STATE>" you've " |
<EXTENDED_NAME_STATE>" right " |
<EXTENDED_NAME_STATE>" first " |
<EXTENDED_NAME_STATE>[ ]?"tieline " |
<EXTENDED_NAME_STATE>[ ]?"extension " |
<EXTENDED_NAME_STATE>" {digit} |
<EXTENDED_NAME_STATE>" please " {cout << endl; char *q=yytext+strlen(yytext); while
(q>yytext) unput(*--q); number_type=none; BEGIN PHONE_NUMBER_STATE;}
<EXTENDED_NAME_STATE>. {ECHO;}

```

```

. {}

```

```

%%

```

```

main()

```

```

{

```

```

    yylex();

```

```

    cout << endl;

```

```

}

```



Attorney Docket No. YOR920010253US1

AF
2182

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): G.G. Zweig et al.
Docket No.: YOR920010253US1
Serial No.: 09/898,289
Filing Date: July 3, 2001
Group: 2182
Examiner: Justin R. Knapp

I hereby certify that this paper is being deposited on this date with the U.S. Postal Service as first class mail addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.	
Signature: <i>Lisa L. Thelma</i>	Date: April 29, 2004

Title: Information Extraction from Documents
with Regular Expression Matching

RECEIVED

MAY 05 2004

TRANSMITTAL LETTER

Technology Center 2100

Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Submitted herewith are the following documents relating to the above-identified patent application:

- (1) Response to Final Office Action; and
- (2) Declaration of Prior Invention Under 37 C.F.R. §1.131 with Exhibit(s) (first copy signed by inventor Geoffrey G. Zweig and second copy signed by inventor Mukund Padmanabhan).

In the event of non-payment or improper payment of a required fee, the Commissioner is authorized to charge or to credit **International Business Machines Corporation Deposit Account No. 50-0510** as required to correct the error.

Respectfully submitted,

Date: April 29, 2004

Robert W. Griffith
Reg. No. 48,956
Attorney for Applicant(s)
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-4547

1. Number	Hits	Search Text	DB	Time stamp
5	5748	710/65 or 710/16 or 705/1	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2004/06/25 12:54
6	3839	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with document\$2) and g06f\$ ipc.	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2004/06/25 12:48
7	1	(flex or lex or perl) with (phrase\$2 with extract\$3)	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2004/06/25 12:48
8	495969	((identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2))	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2004/06/25 12:48
9	61	(710/65 or 710/16 or 705/1) and (((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with document\$2) and g06f\$ ipc.)	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2004/06/25 12:48
10	3512	(710/65 or 710/16 or 705/1) and (((identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2)))	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2004/06/25 12:48
11	14	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with (voice near4 message)) and g06f\$ ipc.	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2004/06/25 12:52
12	0	2002/0194026	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2004/06/25 12:52
13	15769	klein.in.	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2004/06/25 12:53
14	4	klein.in. and 705/2.ccls.	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2004/06/25 12:54
15	0	20020194026.URPN.	USPAT	2004/06/25 12:53
16	7162	710/65 or 710/16 or 705/1 or 705/2	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2004/06/25 12:54
17	495969	(identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2)	USPAT; US-PGPUB; EPO, JPO; DERWENT; IBM_TDB	2004/06/25 12:54

Search History 6/25/04 1:06:26 PM Page 1

CA\APPS\EAST\Workspaces\09898289.wsp

BEST AVAILABLE COPY

18	352465	((identif\$5 or recogniz\$5 or extract\$4) with (data or symbol\$2 or word\$2 same expression\$2))	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 12:55
19	129178	(extract\$4) with (data or symbol\$2 or word\$2 same expression\$2)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 12:55
20	1106	((extract\$4) with (data or symbol\$2 or word\$2 same expression\$2)) and (710/65 or 710/16 or 705/1 or 705/2)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 12:55
21	2559	((extract\$4) with (data or symbol\$2 or word\$2 same expression\$2)) same language\$2	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 12:56
22	70	((extract\$4) with (data or symbol\$2 or word\$2 same expression\$2)) same language\$2 and (710/65 or 710/16 or 705/1 or 705/2)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 13:04
23	11220	707/1 or 707/6 or 707/100	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 13:04
25	8124	(707/1 or 707/6 or 707/100) and ((identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2))	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 13:05
26	259	(707/1 or 707/6 or 707/100) and (((extract\$4) with (data or symbol\$2 or word\$2 same expression\$2)) same language\$2)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 13:05
24	2	(707/1 or 707/6 or 707/100) and (((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with (voice near4 message)) and g06f\$.ipc.)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 13:05
-	17846	(identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2) same document\$2	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 14:33
-	6279	(extract\$4) same (data or symbol\$2 or word\$2 same expression\$2) same document\$2	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/28 13:14
-	3242	(extract\$4) with (data or symbol\$2 or word\$2 same expression\$2) with document\$2	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/28 13:34

-	3330	((extract\$4) with (data or symbol\$2 or expression\$2) with document\$2	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/29 16:10
-	8	((extract\$4) with (data adj2 symbol\$2) with document\$2	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/28 13:17
-	3143	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with document\$2) and g06f\$.ipc.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/25 19:44
-	13	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with (voice near4 message)) and g06f\$.ipc.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/25 19:43
-	8	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with (voice near4 mail)) and g06f\$.ipc.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/28 14:11
-	0	20030050782.URPN.	USPAT	2003/07/28 13:57
-	43	((extract\$4) with ((key adj2 word\$2) or information\$3)) with (voice near4 (messag\$3 or mail\$3)) and g06f\$.ipc.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 14:23
-	276	710/65.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 12:43
-	7	(identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2) same document\$2 and 710/65.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/29 16:09
-	0	((extract\$4) with (data or symbol\$2 or expression\$2) with document\$2 and 710/65.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2003/07/29 16:10
-	14	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with (voice near4 message)) and g06f\$.ipc.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 12:48
-	230	710/16.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 14:31
-	459691	(identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/25 19:43

Search History 6/25/04 1:06:26 PM Page 3

C:\APPS\EAST\Workspaces\09898289.wsp

BEST AVAILABLE COPY

-	135	710/16.ccls. and ((identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2))	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 12:48
-	0	(710/16.ccls. and ((identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2))) same voice	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 15:00
-	4	(flex or lex or perl) with (expression\$2 with extract\$3)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 15:07
-	0	(flex or lex or perl) with (phrase with extract\$3)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/23 15:07
-	1	(flex or lex or perl) with (phrase\$2 with extract\$3)	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 12:48
-	2448	705/1.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/25 19:42
-	1433	(identif\$5 or recogniz\$5 or extract\$4) same (data or symbol\$2 or word\$2 same expression\$2) and 705/1.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 12:54
-	0	((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with (voice near4 message)) and g06f\$.ipc.) and 705/1.ccls.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/01/25 19:43
-	34	705/1.ccls. and ((extract\$4) with (data or symbol\$2 or word\$2 or expression\$2) with document\$2) and g06f\$.ipc.	USPAT; US-PGPUB; EPO; JPO; DERWENT; IBM_TDB	2004/06/25 12:47



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

NOTICE OF ALLOWANCE AND FEE(S) DUE

7590 07/02/2004
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560

EXAMINER

KNAPP, JUSTIN R

ART UNIT

PAPER NUMBER

2182

DATE MAILED: 07/02/2004

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/898,289	07/03/2001	Geoffrey G. Zweig	YOR920010253US1	2216

TITLE OF INVENTION: INFORMATION EXTRACTION FROM DOCUMENTS WITH REGULAR EXPRESSION MATCHING

APPLN. TYPE	SMALL ENTITY	ISSUE FEE	PUBLICATION FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	NO	\$1330	\$300	\$1630	10/04/2004

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. **PROSECUTION ON THE MERITS IS CLOSED.** THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN **THREE MONTHS** FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. **THIS STATUTORY PERIOD CANNOT BE EXTENDED.** SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE REFLECTS A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE APPLIED IN THIS APPLICATION. THE PTOL-85B (OR AN EQUIVALENT) MUST BE RETURNED WITHIN THIS PERIOD EVEN IF NO FEE IS DUE OR THE APPLICATION WILL BE REGARDED AS ABANDONED.

HOW TO REPLY TO THIS NOTICE:

I. Review the SMALL ENTITY status shown above.

If the SMALL ENTITY is shown as YES, verify your current SMALL ENTITY status:

A. If the status is the same, pay the TOTAL FEE(S) DUE shown above.

B. If the status is changed, pay the PUBLICATION FEE (if required) and twice the amount of the ISSUE FEE shown above and notify the United States Patent and Trademark Office of the change in status, or

If the SMALL ENTITY is shown as NO:

A. Pay TOTAL FEE(S) DUE shown above, or

B. If applicant claimed SMALL ENTITY status before, or is now claiming SMALL ENTITY status, check the box below and enclose the PUBLICATION FEE and 1/2 the ISSUE FEE shown above.

☐ Applicant claims SMALL ENTITY status.
See 37 CFR 1.27.

II. PART B - FEE(S) TRANSMITTAL should be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). Even if the fee(s) have already been paid, Part B - Fee(s) Transmittal should be completed and returned. If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed and an extra copy of the form should be submitted.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Utility patents issuing on applications filed on or after Dec. 12, 1980 may require payment of maintenance fees. It is patentee's responsibility to ensure timely payment of maintenance fees when due.

Page 1 of 3

PTOL-85 (Rev. 11/03) Approved for use through 04/30/2004.

BEST AVAILABLE COPY

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), to: **Mail****Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
(703) 746-4000**or **Fax**

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 4 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Legibly mark-up with any corrections or use Block 1)

7590

07/02/2004

Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being facsimile transmitted to the USPTO, on the date indicated below.

..... (Depositor's name)
..... (Signature)
..... (Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/898,289	07/03/2001	Geoffrey G. Zweig	YOR920010253US1	2216

TITLE OF INVENTION: INFORMATION EXTRACTION FROM DOCUMENTS WITH REGULAR EXPRESSION MATCHING

APPLN. TYPE	SMALL ENTITY	ISSUE FEE	PUBLICATION FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	NO	\$1330	\$300	\$1630	10/04/2004

EXAMINER	ART UNIT	CLASS-SUBCLASS
KNAPP, JUSTIN R	2182	710-016000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47, Rev 03-02 or more recent) attached. Use of a Customer Number is required.

2. For printing on the patent front page, list (1) the names of up to 3 registered patent attorneys or agents OR, alternatively, (2) the name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 _____
2 _____
3 _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. Inclusion of assignee data is only appropriate when an assignment has been previously submitted to the USPTO or is being submitted under separate cover. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE

(B) RESIDENCE: (CITY AND STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent); ☐ individual ☐ corporation or other private group entity ☐ government

4a. The following fee(s) are enclosed:

☐ Issue Fee☐ Publication Fee☐ Advance Order - # of Copies _____

4b. Payment of Fee(s):

☐ A check in the amount of the fee(s) is enclosed.☐ Payment by credit card. Form PTO-2038 is attached.☐ The Director is hereby authorized by charge the required fee(s), or credit any overpayment, to Deposit Account Number _____ (enclose an extra copy of this form).

Director for Patents is requested to apply the Issue Fee and Publication Fee (if any) or to re-apply any previously paid issue fee to the application identified above.

(Authorized Signature)

(Date)

NOTE: The Issue Fee and Publication Fee (if required) will not be accepted from anyone other than the applicant, a registered attorney or agent, or the assignee or other party in interest as shown by the records of the United States Patent and Trademark Office.

This collection of information is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, Alexandria, Virginia 22313-1450.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

BEST AVAILABLE COPY

TRANSMIT THIS FORM WITH FEE(S)



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/898,289	07/03/2001	Geoffrey G. Zweig	YOR920010253US1	2216
7590 07/02/2004 Ryan, Mason & Lewis, LLP 90 Forest Avenue Locust Valley, NY 11560			EXAMINER KNAPP, JUSTIN R.	
			ART UNIT 2182	PAPER NUMBER

DATE MAILED: 07/02/2004

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment to date is 329 day(s). If the issue fee is paid on the date that is three months after the mailing date of this notice and the patent issues on the Tuesday before the date that is 28 weeks (six and a half months) after the mailing date of this notice, the Patent Term Adjustment will be 329 day(s).

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) system (<http://pair.uspto.gov>).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (703) 305-1383. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at (703) 305-8283.

Notice of Allowability	Application No.	Applicant(s)	
	09/898,289	ZWEIG ET AL.	
	Examiner	Art Unit	
	Justin Knapp	2182	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to 05/03/04.
2. ☒ The allowed claim(s) is/are 1-25.
3. ☒ The drawings filed on 09/05/01 are accepted by the Examiner.
4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
 - a) ☐ All b) ☐ Some* c) ☐ None of the:
 1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

* Certified copies not received: _____.

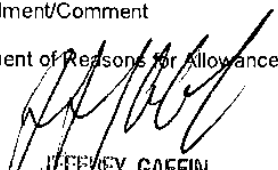
Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.
6. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
 - (a) ☐ including changes required by the Notice of Draftperson's Patent Drawing Review (PTO-948) attached
 - 1) ☐ hereto or 2) ☐ to Paper No./Mail Date _____.
 - (b) ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____.

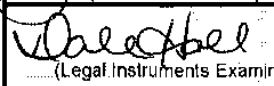
Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).
7. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

- | | |
|---|---|
| 1. ☐ Notice of References Cited (PTO-892) | 5. ☐ Notice of Informal Patent Application (PTO-152) |
| 2. ☐ Notice of Draftperson's Patent Drawing Review (PTO-948) | 6. ☐ Interview Summary (PTO-413),
Paper No./Mail Date _____ |
| 3. ☐ Information Disclosure Statements (PTO-1449 or PTO/SB/08),
Paper No./Mail Date _____ | 7. ☐ Examiner's Amendment/Comment |
| 4. ☐ Examiner's Comment Regarding Requirement for Deposit
of Biological Material | 8. ☐ Examiner's Statement of Reasons for Allowance |
| | 9. ☐ Other _____ |


JEFFREY GAFFIN
 SUPERVISOR/PATENT EXAMINER
 TECHNOLOGY CENTER 2100

Issue Classification 	Application No.	Applicant(s)	
	09/898,289	ZWEIG ET AL.	
	Examiner	Art Unit	
	Justin Knapp	2182	

ISSUE CLASSIFICATION											
ORIGINAL					CROSS REFERENCE(S)						
CLASS	SUBCLASS				CLASS	SUBCLASS (ONE SUBCLASS PER BLOCK)					
710	16				710	65					
INTERNATIONAL CLASSIFICATION					707	1	6	100			
g	0	6	F	3/00	705	1	2				
				/							
				/							
				/							
				/							
 6/25/04 (Assistant Examiner) (Date)					JEFFREY GAFFIN SUPERVISORY PATENT EXAMINER TECHNOLOGY CENTER 2100  6/25/04 (Primary Examiner) (Date)					Total Claims Allowed: 25 O.G. Print Claim(s) 1	O.G. Print Fig. 1
 6/29/04 (Legal Instruments Examiner) (Date)											

☐ Claims renumbered in the same order as presented by applicant		☐ CPA		☐ T.D.		☐ R.1.47	
Final	Original	Final	Original	Final	Original	Final	Original
1	1		31		61		91
2	2		32		62		92
3	3		33		63		93
4	4		34		64		94
5	5		35		65		95
6	6		36		66		96
7	7		37		67		97
8	8		38		68		98
9	9		39		69		99
10	10		40		70		100
11	11		41		71		101
12	12		42		72		102
13	13		43		73		103
14	14		44		74		104
15	15		45		75		105
16	16		46		76		106
17	17		47		77		107
18	18		48		78		108
19	19		49		79		109
20	20		50		80		110
21	21		51		81		111
22	22		52		82		112
23	23		53		83		113
24	24		54		84		114
25	25		55		85		115
	26		56		86		116
	27		57		87		117
	28		58		88		118
	29		59		89		119
	30		60		90		120
							121
							122
							123
							124
							125
							126
							127
							128
							129
							130
							131
							132
							133
							134
							135
							136
							137
							138
							139
							140
							141
							142
							143
							144
							145
							146
							147
							148
							149
							150
							151
							152
							153
							154
							155
							156
							157
							158
							159
							160
							161
							162
							163
							164
							165
							166
							167
							168
							169
							170
							171
							172
							173
							174
							175
							176
							177
							178
							179
							180
							181
							182
							183
							184
							185
							186
							187
							188
							189
							190
							191
							192
							193
							194
							195
							196
							197
							198
							199
							200
							201
							202
							203
							204
							205
							206
							207
							208
							209
							210

U.S. Patent and Trademark Office

BEST AVAILABLE COPY

Part of Paper No. 062504



Attorney Docket No. YOR920010253US1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): G.G. Zweig et al.
Docket No.: YOR920010253US1
Serial No.: 09/898,289
Filing Date: July 3, 2001
Group: 2182
Examiner: Justin R. Knapp

I hereby certify that this paper is being deposited on this date with the U.S. Postal Service as first class mail addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Signature: Lisa L. Chulpa Date: April 29, 2004

RECEIVED

MAY 05 2004

Technology Center 2100

Title: Information Extraction from Documents
with Regular Expression Matching

RESPONSE TO FINAL OFFICE ACTION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In response to the outstanding final Office Action dated January 29, 2004, please consider the following remarks:

Please
Enter
[Signature]
6/28/04

BEST AVAILABLE COPY

Search Notes 	Application No.		Applicant(s)	
	09/898,289		ZWEIG ET AL.	
	Examiner		Art Unit	
	Justin Knapp		2182	

SEARCHED			
Class	Subclass	Date	Examiner
710	65	7/28/2003	JK
710	16	↓	↓
707	1	↓	↓
707	6	↓	↓
707	100	↓	↓
705	1	↓	↓
705	2	↓	↓
updated		6/24-25/04	↓

INTERFERENCE SEARCHED			
Class	Subclass	Date	Examiner
710	16, 65	6/24/04	JK
707	1,6,100	↓	↓

SEARCH NOTES (INCLUDING SEARCH STRATEGY)		
	DATE	EXMR
East Search	7/28/2003	JK
Updated	6/23-25/04	↓

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), to: Mail

Mail Stop ISSUE FEE
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
(703) 746-4000

or Fax

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 4 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Legibly mark-up with any corrections or use Block 1)

7590

07/02/2004

Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560



Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being facsimile transmitted to the USPTO, on the date indicated below.

(Depositor's name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/898,289	07/03/2001	Geoffrey G. Zweig	YOR920010253US1	2216

TITLE OF INVENTION: INFORMATION EXTRACTION FROM DOCUMENTS WITH REGULAR EXPRESSION MATCHING

APPLN. TYPE	SMALL ENTITY	ISSUE FEE	PUBLICATION FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	NO	\$1330	\$300	\$1630	10/04/2004

EXAMINER	ART UNIT	CLASS-SUBCLASS
KNAPP, JUSTIN R	2182	710-016000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.563).

- ☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.
- ☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-02 or more recent) attached. Use of a Customer Number is required.

2. For printing on the patent front page, list (1) the names of up to 3 registered patent attorneys or agents OR, alternatively, (2) the name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1. Thu Ann Dang

2. Ryan, Mason & Lewis, LLP

3. _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. Inclusion of assignee data is only appropriate when an assignment has been previously submitted to the USPTO or is being submitted under separate cover. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE

(B) RESIDENCE: (CITY and STATE OR COUNTRY)

International Business Machines Corporation

Armonk, NY

Please check the appropriate assignee category or categories (will not be printed on the patent); ☐ individual ☒ corporation or other private group entity ☐ government

4a. The following fee(s) are enclosed:

- ☒ Issue Fee
- ☒ Publication Fee
- ☐ Advance Order - # of Copies _____

4b. Payment of Fee(s):

- ☐ A check in the amount of the fee(s) is enclosed.
- ☐ Payment by credit card. Form PTO-2038 is attached.
- ☒ The Director is hereby authorized to charge the required fee(s), or credit any overpayment, to Deposit Account Number 50-0510 (enclose an extra copy of this form).

Director for Patents is requested to apply the Issue Fee and Publication Fee (if any) or to re-apply any previously paid issue fee to the application identified above.

(Authorized Signature) William E. Lewis (Date) July 15, 2004

William E. Lewis, Reg. No. 39,274

NOTE: The Issue Fee and Publication Fee (if required) will not be accepted from anyone other than the applicant; a registered attorney or agent; or the assignee or other party in interest as shown by the records of the United States Patent and Trademark Office.

This collection of information is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, Alexandria, Virginia 22313-1450.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

09/29/2004 TBESHAHE 00000150 500510 09698289

01 FC:1501 1330.00 DA
02 FC:1504 300.00 DA

TRANSMIT THIS FORM WITH FEE(S)

PTOL-85 (Rev. 11/03) Approved for use through 04/30/2004.

OMB 0651-0033

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

AO 120 (Rev. 08/16)

Mail Stop 8 TO: Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Eastern District of Texas, Marshall Division on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292):

DOCKET NO. 2:21-cv-00441	DATE FILED 11/30/2021	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT TREND MICRO INCORPORATED
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 6,842,796	1/11/2005	Taasera Licensing LLC
2 7,673,137	3/2/2010	Taasera Licensing LLC
3 8,327,441	12/4/2012	Taasera Licensing LLC
4 8,850,517	9/30/2014	Taasera Licensing LLC
5 8,955,038	2/10/2015	Taasera Licensing LLC

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGMENT

CLERK	(BY) DEPUTY CLERK	DATE
-------	-------------------	------

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
 Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

PAGE 1 of 2

AO 120 (Rev. 08/16)

Mail Stop 8 TO: Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Eastern District of Texas, Marshall Division on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292):

DOCKET NO. 2:21-cv-00441	DATE FILED 11/30/2021	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT TREND MICRO INCORPORATED
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 8,990,948	3/24/2015	Taasera Licensing LLC
2 9,092,616	7/28/2015	Taasera Licensing LLC
3 9,608,997	3/28/2017	Taasera Licensing LLC
4 9,923,918	3/20/2018	Taasera Licensing LLC
5		

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY	
	☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGMENT

CLERK	(BY) DEPUTY CLERK	DATE
-------	-------------------	------

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
 Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

AO 120 (Rev. 08/10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court _____ for the Eastern District of Texas, Marshall Division _____ on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.);

DOCKET NO. 2:21-cv-00441	DATE FILED 11/30/2021	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT TREND MICRO INCORPORATED
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 6,842,796	1/11/2005	Taasera Licensing LLC
2 7,673,137	3/2/2010	Taasera Licensing LLC
3 8,327,441	12/4/2012	Taasera Licensing LLC
4 8,850,517	9/30/2014	Taasera Licensing LLC
5 8,955,038	2/10/2015	Taasera Licensing LLC

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT		
CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
 Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

PAGE 1 of 2

AO 120 (Rev. 08/10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Eastern District of Texas, Marshall Division on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.);

DOCKET NO. 2:21-cv-00441	DATE FILED 11/30/2021	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT TREND MICRO INCORPORATED
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 8,990,948	3/24/2015	Taasera Licensing LLC
2 9,092,616	7/28/2015	Taasera Licensing LLC
3 9,608,997	3/28/2017	Taasera Licensing LLC
4 9,923,918	3/20/2018	Taasera Licensing LLC
5		

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY	
	☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT		
CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
 Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

PAGE 2 of 2

AO 120 (Rev. 08/10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Eastern District of Texas, Marshall Division on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.);

DOCKET NO. 2:22-cv-00062	DATE FILED 2/25/2022	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT PALO ALTO NETWORKS, INC.
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 6,842,796	1/11/2005	Taasera Licensing LLC
2 7,673,137	3/2/2010	Taasera Licensing LLC
3 8,127,356	2/28/2012	Taasera Licensing LLC
4 8,327,441	12/4/2012	Taasera Licensing LLC
5 8,850,517	9/30/2014	Taasera Licensing LLC

In the above entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY		
	☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading		
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK	
1			
2			
3			
4			
5			

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE
-------	-------------------	------

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
 Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

AO 120 (Rev. 08/10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Eastern District of Texas, Marshall Division on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.):

DOCKET NO. 2:22-cv-00062	DATE FILED 2/25/2022	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT PALO ALTO NETWORKS, INC.
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 8,955,038	2/10/2015	Taasera Licensing LLC
2 8,990,948	3/24/2015	Taasera Licensing LLC
3 9,092,616	7/28/2015	Taasera Licensing LLC
4 9,923,918	3/20/2018	Taasera Licensing LLC
5		

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT		
CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
 Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

AO 100 (Rev. 08/19)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	--

In Compliance with 35 U.S.C. § 290 and/or 35 U.S.C. § 1116 you are hereby advised that a court action has been
 filed in the U.S. District Court **for the Eastern District of Texas, Marshall Division** on the following

☐ Trademarks or ☒ Patents (☐ the patent action involves 35 U.S.C. § 292.)

DOCKET NO. 2:22-cv-00063	DATE FILED 2/25/2022	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT CHECK POINT SOFTWARE TECHNOLOGIES LTD.
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 6,842,796	1/11/2005	Taasera Licensing LLC
2 8,327,441	12/4/2012	Taasera Licensing LLC
3 8,955,038	2/10/2015	Taasera Licensing LLC
4 8,990,948	3/24/2015	Taasera Licensing LLC
5 9,092,616	7/28/2015	Taasera Licensing LLC

In the above-captioned case, the following patent(s)/trademark(s) have been included:

DATE INCLUDED	INCLUDED BY	
	☐ Assignment ☐ Answer ☐ Cross Petition ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above-captioned case, the following decision has been rendered or judgment issued:

DECISION/JUDGMENT

CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director. Copy 3—Upon termination of action, mail this copy to Director.
 Copy 2—Upon filing document adding patent(s), mail this copy to Director. Copy 4—Case file copy.

AO 100 (Rev. 08/19)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	--

In Compliance with 35 U.S.C. § 290 and/or 35 U.S.C. § 1116 you are hereby advised that a court action has been
 filed in the U.S. District Court **for the Eastern District of Texas, Marshall Division** on the following

☐ Trademarks or ☒ Patents (☐ the patent action involves 35 U.S.C. § 292.)

DOCKET NO. 2:22-cv-00063	DATE FILED 2/25/2022	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT CHECK POINT SOFTWARE TECHNOLOGIES LTD.
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 9,608,997	3/28/2017	Taasera Licensing LLC
2 9,923,918	3/20/2018	Taasera Licensing LLC
3		
4		
5		

In the above-captioned case, the following patent(s)/trademark(s) have been included:

DATE INCLUDED	INCLUDED BY	
	☐ Assignment ☐ Answer ☐ Cross Petition ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above-captioned case, the following decision has been rendered or judgment issued:

DECISION/JUDGMENT		
CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director. Copy 3—Upon termination of action, mail this copy to Director.
 Copy 2—Upon filing document adding patent(s), mail this copy to Director. Copy 4—Case file copy.

AO 120 (Rev. 08/10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Eastern District of Texas, Marshall Division on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.);

DOCKET NO. 2:22-cv-00062	DATE FILED 2/25/2022	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT PALO ALTO NETWORKS, INC.
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 6,842,796	1/11/2005	Taasera Licensing LLC
2 7,673,137	3/2/2010	Taasera Licensing LLC
3 8,127,356	2/28/2012	Taasera Licensing LLC
4 8,327,441	12/4/2012	Taasera Licensing LLC
5 8,850,517	9/30/2014	Taasera Licensing LLC

In the above entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK <i>David A. O'Poole</i>	(BY) DEPUTY CLERK nkl	DATE 3/22/22
----------------------------------	--------------------------	-----------------

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

AO 120 (Rev. 08/10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Eastern District of Texas, Marshall Division on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.):

DOCKET NO. 2:22-cv-00062	DATE FILED 2/25/2022	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT PALO ALTO NETWORKS, INC.
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 8,955,038	2/10/2015	Taasera Licensing LLC
2 8,990,948	3/24/2015	Taasera Licensing LLC
3 9,092,616	7/28/2015	Taasera Licensing LLC
4 9,923,918	3/20/2018	Taasera Licensing LLC
5		

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading		
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK	
1			
2			
3			
4			
5			

In the above entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT		
In light of the Notice, which the Court ACCEPTS AND ACKNOWLEDGES, and pursuant to Rule 41(a)(1)(A)(i), all pending claims and causes of action in the above-captioned case are DISMISSED WITHOUT PREJUDICE		
CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

AO 120 (Rev. 08/10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Southern District of New York on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.)

DOCKET NO. 22cv2306	DATE FILED 3/22/2022	U.S. DISTRICT COURT for the Southern District of New York
PLAINTIFF PALO ALTO NETWORKS, INC.		DEFENDANT TAASERA LICENSING LLC, and QUEST PATENT RESEARCH CORPORATION
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 6,842,796	1/11/2005	Taasera Licensing LLC
2 7,673,137	3/2/2010	Taasera Licensing LLC
3 8,127,356	2/28/2012	Taasera Licensing LLC
4 8,327,441	12/4/2012	Taasera Licensing LLC
5 8,850,517	9/30/2014	Taasera Licensing LLC

In the above entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK RUBY J. KRAJICK	(BY) DEPUTY CLERK /S/ V. BRAHIMI	DATE 3/23/2022
--------------------------	-------------------------------------	-------------------

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

AO 120 (Rev. 08/10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Southern District of New York on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.)

DOCKET NO. 22cv2306	DATE FILED 3/22/2022	U.S. DISTRICT COURT for the Southern District of New York
PLAINTIFF PALO ALTO NETWORKS, INC.		DEFENDANT TAASERA LICENSING LLC, and QUEST PATENT RESEARCH CORPORATION
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 8,955,038	2/10/2015	Taasera Licensing LLC
2 8,990,948	3/24/2015	Taasera Licensing LLC
3 9,902,616	7/28/2015	Taasera Licensing LLC
4 9,923,918	3/20/2018	Taasera Licensing LLC
5		

In the above entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK Ruby J. KRAJICK	(BY) DEPUTY CLERK /S/ V. BRAHIMI	DATE 3/23/2022
--------------------------	-------------------------------------	-------------------

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

AO 120 (Rev. 08/10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Eastern District of Texas, Marshall Division on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.)

DOCKET NO. 2:22-cv-00062	DATE FILED 2/25/2022	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT PALO ALTO NETWORKS, INC.
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 8,842,796	1/11/2005	Taasera Licensing LLC
2 7,673,137	3/2/2010	Taasera Licensing LLC
3 8,127,356	2/28/2012	Taasera Licensing LLC
4 8,327,441	12/4/2012	Taasera Licensing LLC
5 8,850,517	9/30/2014	Taasera Licensing LLC

In the above entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY	☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT		
CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

AO 120 (Rev. 08-10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Eastern District of Texas, Marshall Division on the following

☐ Trademarks or ☒ Patents. ☐ the patent action involves 35 U.S.C. § 292.)

DOCKET NO. 2:22-cv-00062	DATE FILED 2/25/2022	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TASERA LICENSING LLC		DEFENDANT PALO ALTO NETWORKS, INC.
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 8,955,038	2/10/2015	Tasera Licensing LLC
2 8,990,948	3/24/2015	Tasera Licensing LLC
3 9,092,616	7/28/2015	Tasera Licensing LLC
4 9,923,918	3/20/2018	Tasera Licensing LLC
5		

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

AO 120 (Rev. 08/10)

Mail Stop 8 TO: Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Eastern District of Texas, Marshall Division on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.);

DOCKET NO. 2:22-cv-00063	DATE FILED 2/25/2022	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT CHECK POINT SOFTWARE TECHNOLOGIES LTD.
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 6,842,796	1/11/2005	Taasera Licensing LLC
2 8,327,441	12/4/2012	Taasera Licensing LLC
3 8,955,038	2/10/2015	Taasera Licensing LLC
4 8,990,948	3/24/2015	Taasera Licensing LLC
5 9,092,616	7/28/2015	Taasera Licensing LLC

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE
-------	-------------------	------

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
 Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

AO 120 (Rev. 08/10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Eastern District of Texas, Marshall Division on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.)

DOCKET NO. 2:22-cv-00063	DATE FILED 2/25/2022	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT CHECK POINT SOFTWARE TECHNOLOGIES LTD.
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 9,608,997	3/28/2017	Taasera Licensing LLC
2 9,923,918	3/20/2018	Taasera Licensing LLC
3		
4		
5		

In the above entitled case, the following patent(s) trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE
-------	-------------------	------

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
 Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

AO 120 (Rev. 08/10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court _____ for the Eastern District of Texas, Marshall Division _____ on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.);

DOCKET NO. 2:22-cv-00063	DATE FILED 2/25/2022	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT CHECK POINT SOFTWARE TECHNOLOGIES LTD.
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 6,842,796	1/11/2005	Taasera Licensing LLC
2 8,327,441	12/4/2012	Taasera Licensing LLC
3 8,955,038	2/10/2015	Taasera Licensing LLC
4 8,990,948	3/24/2015	Taasera Licensing LLC
5 9,092,616	7/28/2015	Taasera Licensing LLC

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy

AO 120 (Rev. 08/10)

TO: Mail Stop 8 Director of the U.S. Patent and Trademark Office P.O. Box 1450 Alexandria, VA 22313-1450	REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK
---	---

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the Eastern District of Texas, Marshall Division on the following

☐ Trademarks or ☒ Patents. (☐ the patent action involves 35 U.S.C. § 292.);

DOCKET NO. 2:22-cv-00063	DATE FILED 2/25/2022	U.S. DISTRICT COURT for the Eastern District of Texas, Marshall Division
PLAINTIFF TAASERA LICENSING LLC		DEFENDANT CHECK POINT SOFTWARE TECHNOLOGIES LTD.
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1 9,608,997	3/28/2017	Taasera Licensing LLC
2 9,923,918	3/20/2018	Taasera Licensing LLC
3		
4		
5		

In the above-entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY	
	☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading	
PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1		
2		
3		
4		
5		

In the above-entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT		
CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director
 Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy