Model for Evaluating the Effectiveness of Search Operations

Kulik Sergey Dmitrievich

National Research Nuclear University MEPhI (Moscow Engineering Physics Institute),
Kashirskoe shosse 31, 115409, Moscow, Russian Federation
Email: sedmik@mail.ru

Abstract. An automated search with human involvement consisting of two stages is given detailed consideration in this paper. In the first stage, a search without direct human involvement is implemented. In the second stage, the search assumes human involvement. Evaluations of the search operations' effectiveness are presented. These operations are implemented for searching one object exclusively among a variety of similar objects. The average number of similar objects recommended for further analysis was used as effectiveness indicator. A set of numerical evaluation criteria for search effectiveness is introduced. The basis of the search block is a pattern recognition algorithm characterized by two probabilities: 1) probability of missing a target, and 2) false alarm probability. An analytical model of the search block was developed. In this paper particular attention is given to the average length of the recommendatory list as an effectiveness indicator. Four properties of this indicator were determined.

Keywords: *analytical model; false alarm probability; operations research; pattern recognition; probability of missing a target; search operations.*

1 Introduction

Different models of operations research [1] make it possible to implement the necessary evaluation of the effectiveness of different operations, including search operations. In practice these operations can be implemented in different fields. For instance, they can be used in criminalistics for searching a fingerprint in a database (DB) of similar fingerprints, in medicine for searching for people with similar illnesses, on the internet for searching factographical data in response to a user's enquiry, in the context of search operations on land and at sea looking for the image of a lost object, such as a ship or a plane, in a database of set images, and other different fields of practical activity. Operations research is an important stage in the evaluation of a search operation's effectiveness and decision-making. In analyzing operations, different tasks should be decided. Some of these tasks have been reviewed in [2-5]. In some cases, search operations are an essential part of a decision making process. Informative search [2,4,6,7] plays a great role in information systems. Effective search is very

Received October 9th, 2014, Revised July 3rd, 2015, Accepted for publication October 13th, 2015.
Copyright © 2015 Published by ITB Journal Publisher, ISSN: 2337-5787, DOI: 10.5614/itbj.ict.res.appl.2015.9.2.5

important and is sometimes even critical for information systems where the informative search is implemented.

In this paper, a two-stage automated search, analogous to factographical search [4], is considered in detail. In the first stage, the search is implemented, as a rule without human involvement. There are no laborious manual operations. The search in this stage is usually implemented with the help of software and hardware means. These means allow obtaining preliminary search results, for example in the form of a recommendatory list (RL). This RL is formed in accordance with the practical application of the search means. Let us suppose that necessary pattern recognition algorithms [8,9], for example neural network algorithms [10-12], are used in this stage. In the second stage, the human operator, who is a specialist in the area of application, for instance a criminalist or an analyst of pictures from space, joins the search operation. In this stage the human operator makes the final decisions about the search results.

Many papers have been devoted to search problems, for example [2,4-7], mentioning only some of them. However, the task of evaluating the effectiveness of search operations while taking into consideration the limited capabilities of the human operator is not completely covered. The present paper partly eliminates this disadvantage.

In this paper, we introduce an analytical model to determine the effectiveness of search operations using probability theory, mathematical statistics and mathematical analysis. An analytical model was developed using probability theory. Experimental study of the analytical model and processing of the results was carried out by methods of mathematical statistics. An analytical study was carried out using methods of mathematical analysis.

Generally, the task of searching in a search system [13-15] is to find required information connected with an object that is being looked for. For instance, the task of searching in criminalistics is to find an object that is identical to the enquiry (OII) among archive system objects with the help of their descriptions in the search array and, in case of the object’s detection, to collect necessary information about it. Thus, solving crimes or joining criminal cases in searches for criminalistic purposes can be achieved as a result of an effective search realized by the search block (SB).

Let us introduce a set of numerical evaluation criteria–also called effectiveness indicators–to evaluate search operation effectiveness and to compare different variants of SBs. 3 indicators are used for evaluating the effectiveness of search operations:

V_x – probability of the correct response to the search request;

H_x – average number of comparison operations implemented by the search block;

L_x – average RL length outputted by the search block for making a final decision by the human operator.

In this research human operator effectiveness (the effectiveness of the person who makes the final decisions) was evaluated with the help of the following indicator:

L_{FL} (fixed length) – maximum RL length that the human operator can process without errors (in the simplest case – review).

We note that in practice, the inequality $L_x \leq L_{max}$ should be kept. As a rule, this allows preventing errors when the final decisions about the search results are made. To obtain an evaluation of the chosen criteria we need a fixed search block of the search system. Let us consider this block.

2 Search Block

The search system includes:

Block 1 – the object indexing block;
Block 2 – the search block;
Block 3 – the RL processing block;
Block 5 – the recognition block;
Block 6 – the archive of objects;
Block 4 – the search array, for instance, realized as a special database which includes object descriptions in the form of search object patterns (SOPs).

For example, criminal system data or object detection systems that use images with the help of neural network patterns recognition algorithms can be used as search object.

The request for the search, which we will call a search enquiry pattern (SIP), includes a description with sought object attributes. Information in the SOP and SIP can be misrepresented because of different hindrances (noises) or errors during the object indexation.

It is assumed that there are N objects stored in the archive. Each object of the archive is located in a place that is uniquely determined by a registration number (RN). Only one object can be located in one place. All SOPs are stored in the search array of the database in the form of a consecutive linear list. An

SIP includes a description of the object that can be stored or can be absent in the object's array. It is assumed that a request with probability P_z can be a description of the object that is identical to one of the archive objects, where an SOP with probability β_n is in n-place ($n = 1,2,3,\ldots,N$) of the search array. A search request in the form of an SIP goes to the pattern recognition block (Block 5) where the rule of pairwise comparison of the SIP and SOP is applied (for instance, using a neural network algorithm). This rule takes possible misrepresentations into consideration because of different hindrances (noises) or errors (for example, human operator errors).

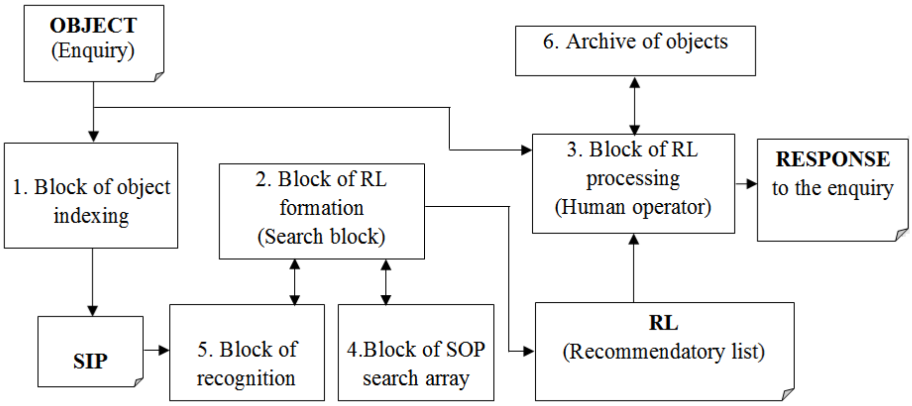

Figure 1 Search object system.

Comparison of the SIP and SOP by the search block is realized using a pattern recognition algorithm and is characterized by probabilities P_1, P_2 [4], where:

1. P_1 – probability of correct comparison of two identical objects based on their descriptions (where (1-P_1) determines the target mission probability);
2. P_2–probability of correct comparison of two non-identical objects based on their descriptions (where (1-P_2) determines the false alarm probability).

If the searching block determines the identity of two objects as a result of pairwise comparison with the help of the recognition block, the SOP registration number is entered into the RL. The total amount of RNs in the RL is L units. During comparison with the SIP, search object patterns appear from each search array one after another for comparison with the search block. When the RL is full or after reviewing the whole search area consisting of N object descriptions in the search array, the search block stops searching. This matches the full search strategy [4]. The result of the search is the RL, which can be empty or can include from 1 to L registration numbers. We consider the length of the RL to be the amount of RNs.

It is important to consider that a human operator can process (browse) only a limited RL with length $L \leq L_{max}$. There is the possibility of changing the length of the RL represented by the search block for the human operator. This possibility allows regulating the quality of the human operator's work. The human operator processes only the objects from the RL, which increases the effectiveness of the search operation.

In some search systems the human operator not only processes the RL. For example, the human operator can also implement object indexation (SIP and SOP formation). During the process of indexation, the human operator can make mistakes (errors). As a result, the SIP or SOP that are included in the search array can have some indexation errors. These errors can lead to missing targets or false alarms.

Generally, objects whose descriptions go to the input of the search block can be described by three groups of attributes. The first and the second groups consist of attributes that can be formally described and can be used in the case of automated object indexation and comparison (for example, by a neural network in the recognition block) with the help of the object descriptions. The first group consists of attributes that are resistant to insignificant object misrepresentation. The second group consists of attributes that are sensitive to insignificant object misrepresentation. An indexation error of these attributes, either by a human operator or occurring automatically (for instance, by a neural network) strongly depends on the level of object misrepresentation. The third group consists of attributes that cannot be formally described and that are used by the human operator (a person who makes a decision) in the case of direct object comparison during the processing of the RL. A scheme of object classification using the descriptions is based on the first group of attributes. With the help of the first group, all object descriptions (SOP and SIP) are divided into 'k' classes, after which the search strategies are implemented. With the help of the second group of attributes, the object recognition (comparison) algorithm using the SIP and SOP descriptions is created. With the help of this algorithm, the search block forms the RL for the human operator. Further, we will assume that there are no attributes that are resistant to insignificant object misrepresentation. Hence, there is no scheme of object classification.

The processing of search enquiries is realized as depicted in Figure 1. Special means (which are not reviewed here) form the stream of enquiries in the form of object enquiries. These object enquiries go to the first block (Block 1) of object indexation, where the object description is compiled in the form of an SIP. In the simplest case, the SIP includes a list of attribute values that will be used in the fifth block (Block 5). As the result of the search, the search block for this SIP finds in the search array the object descriptions (i.e. SOP) that are the most similar to the object enquiry (i.e. SIP). The level of 'similarity' is determined by the comparison (recognition) algorithms of the SOP and SIP. The registration numbers of similar objects (i.e. recommended for further analysis) are entered into the RL. Because of the limited buffer under the RL, the search stops after filling the RL with L registration numbers or after viewing a whole searching area, which includes N units of SOPs in the search array. Furthermore, the formed RL, the enquiry object and the objects-of-storage, whose registered numbers are noted in the RL, are transmitted to the third block (Block 3) for RL processing by the human operator. The human operator makes the final decision (in the output of the search system) by way of analysis (expertise) and comparison of the enquiry object with the objects-of-storage that have been compiled in the RL.

As a result, we can get the response 'Yes'–the same object exists in the system (is identical to the enquiry object) and there is information about this object. We can get the response 'No'–the same object does not exist in the system (is not identical to the enquiry object). The response to a search request is transmitted to the user who has sent the search enquiry.

Let us briefly review the human operator's work on the output of the search system (Figure 1) during the technological operation of the RL processing. The human operator's work during the processing of the preliminary search results in the form of an RL is usually hard, tense and not ideal. This greatly influences the final result of the whole search. The result is the final response to the search enquiry. For example, if the enquiry object is placed towards the end of the RL and the RL is too long, the human operator may stop processing before reaching the end and does not get a correct response to the enquiry. It is important that the human operator's work on the output of the search system takes place at the end of the searching operations. The main task for the human operator is to process the RL and to form the final response to the enquiry. If the RL is too long (consists of many objects that are similar to the enquiry object), the human operator will spend too much time and a lot of effort in processing it. What is more, the human operator will get very tired. Sometimes there are restrictions in the form of deadlines for total processing time (T_{RL}) regardless of the RL length. This means that if the RL has an extended length, the human operator has less time for processing one object of the RL. It is clear that the longer the RL, the more intense and difficult the working process of the human operator (time for reviewing the whole RL is limited by a deadline). As a result, the human operator can make errors because the RL is too long. These errors lead to missing a target (the OII is not given as a response to the enquiry object). Further lengthening of the RL, as a rule, leads to many mistakes (errors) in the human operator's work. As a result, we can see full failure in the processing of an overly long RL.

We assumed that the necessary hardware and software means function without errors and the probability of a correct response to the enquiry is determined by the probability of a correct response to the enquiry by the search block, P_1 and P_2 characteristics of the pattern recognition algorithm (Block 5), and human operator possibilities ($L \le L_{max}$).

This part of the paper demonstrates what kinds of search operations are used in criminalistics. For instance, they can be used for searching a similar gun cartridge case. Suppose there has been a crime and the offender used a pistol (gun). Suppose also that only one cartridge case was found at the crime scene. This cartridge case is placed in the criminalistical collection of cartridge cases (forensic collection of similar cartridge cases) and then a description of the cartridge case is placed in the search array. Suppose there is a suspect. Some gun was confiscated from him, as well as a cartridge case from this gun. This cartridge case needs to be checked against the forensic collection of cartridge cases. There are two kinds of search operations: 1) the search (in a forensic collection–archive of objects) for cartridge cases similar to the cartridge case in question, which are entered as the enquiry object for the search; 2) the search (in a special database–search array) for cartridge case descriptions similar to the cartridge case in question, which are entered as the enquiry object for the search. We focus on the second kind of search operation, which will be evaluated.

3 Analytical Model of the Search Block

Different models are known (for instance [16-19]), but as of yet there has been no appropriate model for the search block. The necessary researches were realized and these enabled some results to be obtained. We will take a set of efficiency indicators and formulas for their evaluation as an analytical model. Analytical formulas were obtained to evaluate the search effectiveness with the help of 3 indicators (V_x, T_x, L_x):

$$V_x = Pz \cdot S1 + (1 - Pz) \cdot S2;$$

$$T_x = \{Pz \cdot W1 + (1 - Pz) \cdot W2\} \cdot t;$$

$$L_x = Pz \cdot F1 + (1 - Pz) \cdot F2;$$

where:

S_1 : probability of the correct response to the enquiry by the search block during the search in the area that includes OII;

S_2 : probability of the correct response to the enquiry by the search block during the search in the area that does not include OII;
W_1 : average number of SIP and SOP pairwise comparisons during the search in the area that includes OII;
W_2 : average number of SIP and SOP pairwise comparisons during the search in the area that does not include OII;
t : average time of one recognition (SIP and SOP comparison); for example, working time of a neural network;
F_1 : average length of the RL during the search in the area that includes OII;
F_2 : average length of the RL during the search in the area which that not include OII.

We note that if $t = 1$, the following expression is correct: $T_x = H_x$.

An analytical model (formula for L_x) was obtained to evaluate the effectiveness of search operations with the help of 3 indicators, F_1, F_2 and P_z. When comparing two possible embodiments of search operations it is best to rely on one that has a minimal value L_x. Suppose that there are 2 embodiments of search operations and their indicators for L_x: $L_x(1)$ and $L_x(2)$. We used the following rules for comparing the metrics to conclude the effectiveness of a search operation:

If $L_x(2) > L_x(1)$, the first embodiment of the search operation is better than the second.

If $L_x(1) > L_x(2)$, the second embodiment of the search operation is better than the first.

If $L_x(1) = L_x(2)$, the first embodiment of the search operation is equivalent to the second.

A similar rule was used for the F_2 indicator.

What is more, if $P_z \approx 0$, the average length of the RL given by the search block for the human operator (person who makes the final decision) is $L_x \approx F_2$.

Let us briefly consider a space of elementary events [20].

To get the V_x evaluation–the probability of the correct response to the enquiry–the Ω space is introduced. The elementary events are responses of the search block to the enquiry. The Ω space of elementary events is divided into 2 disjoint spaces–D and E. The events of the correct response of the search block to the enquiry match the D space. The events of the incorrect response of the search block to the enquiry match the E space. This allows us to calculate the probability of appearance of at least 1 event in the D space. This probability will be taken as the probability of the correct response of the search block to the enquiry. Analogously, we will obtain that V_0 is the probability of the appearance of at least 1 event in the E space. This probability will be taken as the probability of incorrect response of the search block to the enquiry.

The events in the Ω space should be a full group of events. This means that the probability of the appearance of at least 1 event in the Ω space should be equal to 1, i.e. $V_x + V_0 \equiv 1$.

For S_1 probability, the D space can be determined in different ways in accordance with practical applications. Let us consider only one of them. This D space consists of events in which the RN of the identical object is entered into the RL. Moreover, from 1 to (L-1) object registration numbers can be entered into the RL but they are not identical to the enquiry object.

For S_2 probability, the D space can be also determined in different ways in accordance with practical applications. We will consider only one of them. This D space consists of events in which from 1 to L registration numbers that are not identical to the enquiry object are entered into the RL.

Elementary events in the Ω space for L_x are entered analogously.

Following the monograph [20] we denote:

$$\binom{n}{r} = \frac{n\cdot(n-1)\cdot\ldots\cdot(n-r+1)}{1\cdot 2\cdot\ldots\cdot(r-1)\cdot r} = \frac{n!}{r!(n-r)!}, \quad \text{where } n! = n\cdot(n-1)\cdot\ldots\cdot 2\cdot 1.$$

The following analytical expressions for F_1 and F_2 probabilities were obtained to evaluate effectiveness:

$$F1 = F(N, P_1, P_2, L, \beta_n) = F_1 + F_2,$$

$$F_1 = \sum_{n=1}^{L} \beta_n \left\{(1-P_1)\cdot L(N-1, P_2, L) + P_1\left[1 + L(N-1, P_2, L-1)\right]\right\},$$

$$F_2 = \sum_{n=L+1}^{N} \beta_n \left\{\sum_{s=L}^{n-1}\left\{\binom{n-1}{s} P_2^{n-1-s}(1-P_2)^s \cdot L\right\} + \sum_{i=0}^{L-1}\left\{\binom{n-1}{i} P_2^{n-1-i}(1-P_2)^i \cdot Q_i\right\}\right\},$$

$$Q_i = (1-P_1)\left\{L(N-n, P_2, L-i) + i\right\} + P_1\left\{L(N-n, P_2, L-1-i) + 1 + i\right\};$$

$$F2 = L(N, P_2, L) = \sum_{m=0}^{L-1}\left[\binom{N}{m} P_2^{N-m}(1-P_2)^m\, m\right] + \sum_{m=L}^{N}\left[\binom{N}{m} P_2^{N-m}(1-P_2)^m\, L\right].$$

This analytical model was used for two forensic cases: 1) cartridge cases of a gun; 2) forged documents and banknotes. In the first case, the analytical model was used to evaluate the effectiveness of search operations for objects–i.e. cartridge cases of a gun. The model allowed us to estimate the F_2 rate at known values of L, N and P_2. Then we found that S = 5 forensic experts are enough for effective RL processing. In the second case, the analytical model was used to evaluate the effectiveness of search operations for objects–i.e. forged documents and banknotes. The model allowed us to estimate the P_2 rate at known values of L, N and $a < F_2 < b$, where the boundaries of a and b were known.

4 Model Research

It was proved that the functions $F_1 = F(N, P_1, P_2, L, \beta_n)$ and $F_2 = L(N, P_2, L)$ have the following important properties.

Property 1

$$L(N,P_2,L) = \begin{cases} 0 & \text{if } P_2 = 1; \\ L & \text{if } P_2 = 0; \\ N(1-P_2) & \text{if } L = N. \end{cases}$$

Property 2

Function value $L(N, P_2, L)$ does not decrease if N, (1-P_2) or L increases.

Property 3

For $F(N, P_1, P_2, L, \beta_n)$ the following decompositions are correct:

$$F(N, P_1, P_2, L, \beta_n) = P_1\{L(N-1, P_2, L-1) + 1\} + (1-P_1)L(N, P_2, L)$$

$$F(N, P_1, P_2, L, \beta_n) = L(N, P_2, L) + [P_1 + P_2 - 1]\{1 + L(N-1, P_2, L-1) - L(N, P_2, L)\}.$$

Property 4

For any $\delta \geq 0$, if N, L, P_2 are given, parameter E^* exists and with the help of this parameter $|L(N, P_2, L, E^*) - L(N, P_2, L)| \leq \delta$ is implemented, where:

$$L(N,P_2,L,E) = \sum_{m=0}^{L-1}\left[\binom{N}{m} P_2^{N-m}(1-P_2)^m\, m\right] + \sum_{m=L}^{E}\left[\binom{N}{m} P_2^{N-m}(1-P_2)^m\, L\right].$$

For concrete practical applications the necessary concrete E^* evaluation for $L(N, P_2, L, E)$ can be successfully implemented.

For chosen indicators of effectiveness, instrumental software tools for modeling and searching operations research were developed. For some algorithms, their hardware implementation was developed. These means allow us to take the human operator into consideration and evaluate the recognition algorithm's influence on the effectiveness of the search block.

As a result of our researches, it was determined that F_2 changed when L, N, and P_2 changed. A small part of the results from these researches, for different L, N, and P_2, are shown in Tables 1 and 2.

Table 1 Example of Effectiveness, if $N = 1000$ and $P_2 = 0.7$.

Indicators of effectiveness	**Effectiveness indicator values**								
L	20	100	200	250	400	450	480	500	1000
F_2	20	100	200	250	300	300	300	300	300

Table 2 Example of Effectiveness, if $N = 1000$ and $P_2 = 0.75$.

Indicators of effectiveness	**Effectiveness indicator values**									
L	50	200	220	230	265	290	300	500	700	1000
F_2	50	200	220	230	249	250	250	250	250	250

According to Table 1, if $P_2 = 0.7$, $N = 1000$ (N – the number of SOP), and $L = 480$ (L – maximum length of the RL), the length of the RL of the search block is $F_2 = 300$. Analogously, according to Table 2, if $P_2 = 0.75$, $N = 1000$, and $L = 700$, the length of the RL of the search block is $F_2 = 250$. This result is completely coordinated with Property 1:

$$\left|L(N, P_2, L) \approx N(1 - P_2)\right| \quad \text{if} \quad L \approx N$$

or

$$\left|L(N, P_2, L) \approx 1000(1 - 0.7) = 300\right| \quad \text{if} \quad L \approx 480,$$
$$\left|L(N, P_2, L) \approx 1000(1 - 0.75) = 250\right| \quad \text{if} \quad L \approx 700.$$

During the investigation of the average RL length, particular attention was given to the function $L(N, P_2, L)$ because, in accordance with Property 3, we can get the function $F_1 = F(N, \beta P_1) P_2, L, {}_n$ through this function.

Analysis of the dependence $F_2 = L(N, P_2, L)$ on L and N, if the value of P_2 is fixed, shows the following.

1. If $L = N \rightarrow L(N, P_2, L) \approx N \cdot (1 - P_2)$ (this result is completely coordinated with Property 1).
2. If $P_2 \approx 1 \rightarrow L(N,P_2,L) \approx 0$ (this result is completely coordinated with Property 1).
3. If $P_2 \approx 0 \rightarrow F_2 = L(N,P_2,L)$ dependence on L is close to a straight line with a slope of 45º, i.e. $L(N, P_2, L) \approx L$ (this result is completely coordinated with Property 1).
4. If P_2 increases or L decreases $\rightarrow$ function $L(N, P_2, L)$ decreases (this result does not contradict Property 2).

Analysis of the dependence $F_2 = L(N,P_2,L)$ on L and N, if the values of P_2 and L are fixed, shows the following.

1. If N increases $\rightarrow$ the average length of the RL increases as well (this fact is coordinated with Property 2).
2. If $N \rightarrow \infty \rightarrow$ the value of function $L(N,P_2,L)$ tends to L.

The results of the numerical researches completely confirmed the findings of the theoretical researches for the properties of the function $L(N,P_2,L)$. The implemented research allowed creation of a model for evaluating the average RL length and suggested an approach for the acceleration of these calculations (Property 4) if the accuracy of the calculations was set.

5 Experimental Test for the Model

Because F_1 can be obtained through F_2, the received evaluations were experimentally tested for the F_2 indicator of effectiveness. To implement the experimental research, special computer programs were successfully created and used. The experiment confirmed the received evaluation of the F_2 indicator. Several results of this important experimental test can be seen in Table 3.

Table 3 Experimental Test.

Deviation from F_2 (%)	**Proportions of observation cases (%)**
± 5	82
±10	91
±15	96
±20	98

Table 3 displays the proportions of observation cases (%) of deviation value, which was experimentally measured during the theoretical calculation of F_2. We can see that the relative difference of theoretical values of effectiveness indicator F_2 and experimentally measured values for real data did not exceed 15%. We can see this in 96% of cases. Thus, it can be assumed that generally (in 80% of cases), calculation results are equal to F_2 acceptable practical accuracy (about 5%).

6 Practical Usage of the Model

Two stages of evaluating the effectiveness of the human operator and the search block were implemented for successful usage of the created analytical model.

In the first stage, evaluation of the operator's experimental effectiveness was implemented. To do the research, special software was developed, simulating a real search system working in RL view mode, operated by an expert getting the object from the RL and the object enquiry. A decision about their identity was made by the expert with the help of pressing one of two keys. One key was equal to the expert's decision about the object's identity, the other key was equal to the expert's decision about their non-identity. Moreover, there was an opportunity to automatically fix a case of failure when the expert had insufficient time to take a decision and fix objects that were not processed. In the first series of experiments, each of N_φ objects was presented for an equal period τ. In the second series of experiments, the total time (T_{RL}) necessary for viewing N_φ objects (RL processing with N_φ length) was fixed. All experts' responses were automatically fixed and checked. The probability of the expert's work being correct was evaluated in the experiments as $P_E = (1 - N_E/N\varphi)$, where N_E represents the total number of objects viewed by the expert.

The results of experiments with the expert's work on the output of the search system during RL processing are presented in Figure 2. In this figure we can see the dependency of the P_E probability on length L of the RL. For 7 fixed RL length values (L), the experimental values for evaluation of the frequency probability P_E for the expert were obtained, and piecewise linear approximation of missing values was implemented, as shown in Figure 2. This approximation means the following: if two of the nearest experimentally measured values of the P_E probability are known, all other intermediate values between two experimental values are accepted as identical and equal to one of two of these measured values that is the worst (in our case, the lowest) of these two values. This allows the maximum length of the RL to be chosen reasonably with the help of such graphs in future. This maximum length guarantees the provision of a preset mode for the expert's work.

In the first series of experiments, N_{φ} = 1000 and τ accepts discrete values. The results showed that, when the time used for processing one object was fixed, the probability of the expert's work being correct decreased when the length of the RL increased.

In the second series of experiments, T_{PL} was set but was not changed, and the amount of objects N_{φ} took values from 1 to 1000. The expert mode was the following: the whole RL with length N_{φ} was presented to the expert. The total time T_{RL} for RL processing was limited. When time T_{RL} ran out, objects for which the expert had not made a decision were fixed as errors. All responses of the expert were analyzed and all the expert's errors were counted. The results of the second series of experiments confirmed the results of the first series of experiments. When the time necessary for processing the whole RL was fixed, the probability of the expert's work being correct decreased when the length of the RL increased. Note that the results of the individual experiments with real experts and a real object archive confirmed the results previously obtained. (These experiments were analogous to those that were done in the second series but without using the software simulating the search system in RL processing mode.) Moreover, an evaluation of the average time for processing one object without errors was obtained.

The accounting of real experts' characteristics (and modes of their work) can be realized with the help of Figure 2 in the following way: for probability set P_0 with the help of the graph (Figure 2), $L(i)_{max}$ (maximum length of the RL for expert i) and L_{max} (maximum RL length considering all experts):

$$L_{max} = \min_{i \in J} \left\{ L(i)_{max} \right\} .$$

Furthermore, L_x (the average length of the RL given to the expert) is determined with the help of modeling, for example, with the help of the analytical model. If $L_{max} >> L(i)_{max}$, expert i works portions on the average $L(i)_{max}$ objects, or some experts process the RL, the length of which is equal to L_x. After each portion the expert has a rest, restoring his effectiveness. If $L_{max} = L(i)_{max}$, expert i on the average has a rest after processing the succeeding RL.

Another way of accounting is also possible. After defining L_{max} (maximum length of the RL) for T_{RL} (set time) and P_E (probability) with the help of Figure 2, the modeling of the search system is implemented for providing its effectiveness by way of changing or selecting variable parameters–for example, by changing L. N is the number of SOP notations in the search array. So, the possible range of L, $1 \le L \le N$ in this case, can be reduced greatly to $1 \le L \le L_{max} \le N$ (if L_{max} is known). Thus, if the effectiveness of the search block is provided by way of selecting variable parameters (in the case of

$1 \leq L \leq L_{max} \leq N$), the characteristics of the expert will be simultaneously considered (the length of the RL is not more than L_{max} and will be given to the expert). The first way of characteristics accounting was chosen for an experiment with a real expert, and it was assumed that $L_{max} = 25$.

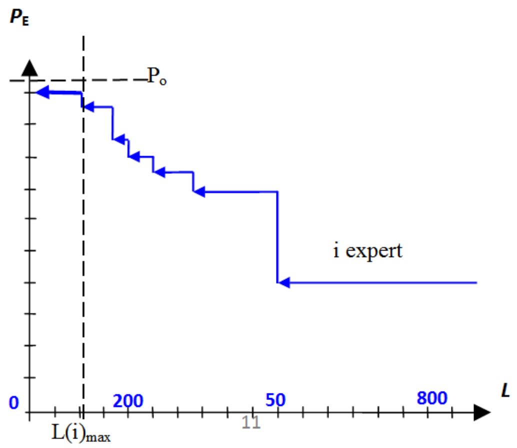

Figure 2 Dependency between P_E and L.

In the second stage, the developed analytical model was practically implemented for evaluation L_x of the process of a criminologist working in a police department (for example, in the process of working with rifle cartridge cases). The effectiveness of an automated search for the criminalistical collection of cartridge cases that were confiscated from unsolved crimes scenes was obtained. A cartridge case was the object of enquiry for the search. Searching means were implemented and the search for cartridge case descriptions similar to those of the cartridge case in question were input as the enquiry for the search.

It was found that $P_z \approx 0$. Hence, we can suppose that $L_x \approx F_2$. As a result, the search means gave the RL, which included indication numbers of similar cartridge cases from 0 to L from the collection (from the object archive). Furthermore, this list was processed by one or several experts. For the search block the evaluation of the average RL length (L_x) with the numbers of cartridge cases that were similar to the enquiry object, was successfully implemented.

In Table 4 we can see the necessary evaluation of the recognition algorithm for $P_2 = 0.87$ with the help of a collection consisting of 950 cartridge cases.

Table 4 Evaluation of Effectiveness for $N = 950$, $P_2 = 0.87$.

L	L_x
1	1
5	5
30	30
70	70
93	92.99
98	97.98
104	103.89
108	107.72
110	109.56
115	113.83
117	115.35
120	117.40
122	118.58
126	120.48
130	121.80
135	122.77
140	123.23
145	123.41
150	**123.48**
155	123.5
160	123.5
161…950	123.5

In Table 4 we can see that if $L > 160$, the specific saturation of L_x for the search block occurred (this is not expedient to increasing parameter L over 140; because saturation occurred, L_x did not increase). Table 4 shows that if $L = 150$ the average length of the RL was $L_x \approx 123.5$

In Figure 3 we can see the correlation (dependency) between L_x and L (L_x depends on L). In this scheme we can select 3 sectors:

1. I $\rightarrow$ almost linear increasing correlation (dependency), $L_x = L$;
2. II $\rightarrow$ non-linear increasing correlation (dependency), $L_x = f(L)$;
3. III $\rightarrow$ almost linear non-increasing correlation (dependency), $L_x = N(1\text{-}P_2)$.

This result is completely coordinated with Properties 1, 2.

Knowing the characteristics of each expert (for example, average time for processing one cartridge case from the RL and the L_x parameter evaluation in the developed model), the necessary number of experts for this RL processing was evaluated. In the first stage, it was determined that $L_{max} = 25$. Following the monograph [21], let's introduce some designations:

$\lceil \cdot \rceil$ – the smallest integer greater than or equal to a number.

Then the rough evaluation necessary for the expert is:

$$S = \left\lceil \frac{L_x}{L_{\max}} \right\rceil \approx \left\lceil \frac{123.5}{25} \right\rceil = \lceil 4.94 \rceil = 5,$$

where $\lceil y \rceil$ – the smallest integer such that $y \le \lceil y \rceil$. As a result, we find that $S = 5$ experts are enough for effective RL processing.

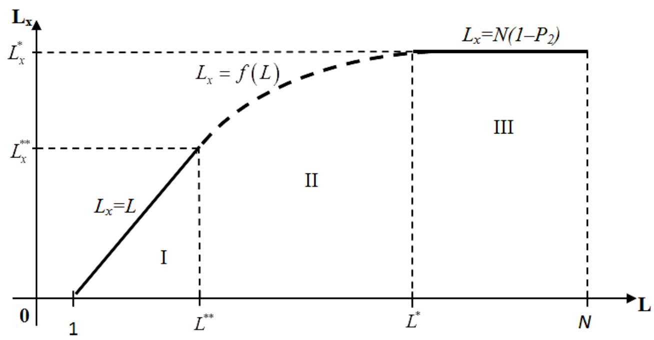

Figure 3 Dependence $L_x \approx L(N,P_2,L)$, if $P_z \approx 0$.

7 Conclusion

As the result of this research, an analytical model allowing the evaluation of the average RL length was developed. Investigation of this model was implemented. Validation of the model (check of the model's adequacy) with the help of experimental data was implemented as well. The properties of an important indicator of effectiveness–the average RL length given by the search block for the human operator–were explored. This allowed a reasonable analysis of the effectiveness of the search implemented by the search block, which is particularly important for providing information safety in criminalistics.

The necessary software, allowing the evaluation of a search's effectiveness before realization of search operations, was created for developers of search blocks. The main characteristics of the human operator and the pattern recognition algorithm were taken into consideration. The obtained results can be used by support systems for making decisions in criminalistics.

We plan to obtain and explore two remaining evaluations of effectiveness in the future. They are:

1. $V_x \rightarrow$ probability of the correct response to the enquiry of the search;
2. $H_x \rightarrow$ average number of comparison operations implemented by the search block.

For these indicators of effectiveness, formulas needed for the evaluation of these indicators, will be offered. The validation of the analytical model (check of adequacy of the analytical model) with the help of two of these indicators will be controlled. Implementation of necessary experimental investigations of this model and the real search block is planned. It is expected that usage of this model will provide positive results during the realization of information search systems.

8 Nomenclature

DB : database
RL : recommendatory list
OII : object that is identical to the enquiry
SB : search block
SOP : search object pattern
SIP : search enquiry pattern
RN : registration number

References

[1] Taha, H.A., *Operations Research: An Introduction* (8th Edition). Upper Saddle River, New Jersey 07458, Pearson Prentice Hall, 2007, p. 813.

[2] Nakamura, K. & Iwai, S., *Topological Fuzzy Sets as a Quantitative Description of Analogical Inference and Its Application to Question-Answering Systems for Information Retrieval*, IEEE Transactions on Systems, Man and Cybernetics, **12**(2), pp. 193-204, March 1982.

[3] Wald, A., *Sequential Analysis* (Second Printing, November, 1948). Mineola, N.Y., Dover Publications, p. 212, 2004.

[4] Kulik, S.D., *Algorithms for Pattern Recognition and Simulation of Automated Factographic Information Retrieval Systems*, Neuro-computers: Design and Application, **9-10**, pp. 115-127, 2002.

[5] Ko, J., Si, L., Nyberg, E. & Mitamura, T., *Probabilistic Models for Answer-Ranking in Multilingual Question-Answering*, ACM Transactions on Information Systems (TOIS), **28**(3), Article 16, p. 37, 2010.

[6] Mateescu G., Sosonkina M. & Thompson P., *A New Model for Probabilistic Information Retrieval on the Web*, Workshop on Web Analytics, J. Ghosh & J. Srivastava, eds., Arlington, Second SIAM International Conference on Data Mining, VA, pp. 17-27, 2002.

[7] Glitho R.H., Olougouna E. & Pierre S., *Mobile Agents and Their Use for Information Retrieval: A Brief Overview and an Elaborate Case Study*, Network IEEE, **16**(1), pp. 34-41, 2002.

[8] Fukunaga, K., *Introduction to Statistical Pattern Recognition* (2nd edition), Elsevier Academic Press, San Diego, San Francisco, New York, Boston, London, Sydney, Tokyo, p. 592, 1990.

[9] Huang H. & He H., *Super-Resolution Method for Face Recognition Using Nonlinear Mappings on Coherent Features*, IEEE Transactions on Neural Networks, **22**(1), pp. 121-130, 2011.

[10] Galushkin, A.I., *Neural Networks Theory*, Springer Berlin Heidelberg New York, p. 396, 2007.

[11] Haykin, S., *Neural Networks–A Comprehensive Foundation, Second Edition*, Pearson Education, Inc., 1999 (reprint), 2005.

[12] Kruglov I.A., Mishulina, O.A. & Bakirov, M.B., *Quantile-based Decision-Making Rule of the Neural Networks Committee for Ill-Posed Approximation Problems*, Neurocomputing, **96**, pp. 74-82, 2012.

[13] Vasconcelos, N., *Minimum Probability of Error Image Retrieval*, IEEE Transactions on Signal Processing, **52**(8), pp. 2322-2336, August, 2004.

[14] Gui, C., Liu, J., Xu, C. & Lu, H., *Web Image Retrieval via Learning Semantics of Query Image*, ICME'09 Proceedings of the 2009 IEEE international conference on Multimedia and Expo (June 28 2009-July 3 2009), IEEE Press Piscataway, NJ, USA, pp. 1476-1479, 2009.

[15] Costa Pereira, J., Coviello, E., Doyle, G., Rasiwasia, N., Lanckriet, G., Levy, R. & Vasconcelos, N., *On the Role of Correlation and Abstraction in Cross-Modal Multimedia Retrieval*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **36**(3), pp. 521-535, 2014.

[16] Blei, D., *Probabilistic Topic Models*, Communications of the ACM, **55**, (4), pp. 77-84, 2012.

[17] Miotto, R. & Orio N., *A Probabilistic Model to Combine Tags and Acoustic Similarity for Music Retrieval*, ACM Transactions on Information Systems (TOIS), **30**(2), Article 8, p. 29, 2012.

[18] Stathopoulos, V. & Jose, J.M., *Bayesian Probabilistic Models for Image Retrieval*, Journal of Machine Learning Research (JMLR): Workshop and Conference Proceedings, **17**, from 2nd Workshop on Applications of Pattern Analysis (WAPA), 19-21 October, 2011, CIEM, Castro Urdiales, Spain, pp. 41-47, 2011.

[19] Losee, R.M. & Church Jr., L., *Information Retrieval with Distributed Databases: Analytic Models of Performance*, IEEE Transactions on parallel and distributed systems, **15**(1), pp. 8-27, 2004.

[20] Feller, W. *An Introduction to Probability Theory and its Applications,* Vol.1, 3rd Ed., John Wiley & Sons. New York, p. 509, 1968.

[21] Imre S. & Balazs, F., *Quantum Computing and Communications: An Engineering Approach*, Chichester, West Sussex, England: John Wiley & Sons, p. 283, 2005.