Int. J. Services Sciences, Vol. 4, No. 1, 2011

DBS-PSI: a new paradigm of database search

Roman Statnikov

Department of Information Sciences,
Naval Postgraduate School,
589 Dyer Rd., Root Hall, 203E, Monterey, CA, USA
and
Mechanical Engineering Research Institute,
Russian Academy of Sciences,
M. Kharitonievsky per., 4, Moscow, 101830, Russia
E-mail: rstatnik@nps.edu

Alex Bordetsky

Department of Information Sciences,
Naval Postgraduate School,
589 Dyer Rd., Root Hall, 225, Monterey, CA, USA
E-mail: abordets@nps.edu

Josef Matusov

Mechanical Engineering Research Institute,
Russian Academy of Sciences,
M. Kharitonievsky per., 4, Moscow, 101830, Russia
E-mail: v_astashev@mail.ru

Alexander Statnikov*

Center for Health Informatics and Bioinformatics,
New York University Langone Medical Center,
333 E38th Street, 6th Floor,
New York, NY 10016, USA
E-mail: Alexander.Statnikov@med.nyu.edu
*Corresponding author

Abstract: The advent of the World Wide Web made search engines the most essential component of our everyday life. However, the analysis of information provided by current search engines often presents a significant challenge to the client. This is to a large extent because the client has to deal with many alternatives (solutions) described by contradictory criteria, when selecting the most preferable (optimal) solutions. Furthermore, criteria constraints cannot be defined a priori and have to be defined interactively in the process of a dialog of the client with computer. In such situations, construction of the feasible solution set has a fundamental value. In this paper, we propose a new methodology for systematically constructing the feasible solution set for database search. This allows to significantly improving the quality of search results.

Keywords: database search; search engine; contradictory criteria; criteria constraints; feasible solution set; Pareto optimal set; parameter space investigation method; PSI method; analysis tables.

Reference to this paper should be made as follows: Statnikov, R., Bordetsky, A., Matusov, J. and Statnikov, A. (2011) 'DBS-PSI: a new paradigm of database search', *Int. J. Services Sciences*, Vol. 4, No. 1, pp.1–13.

Biographical notes: Statnikov received his PhD from Moscow State University of Railways in 1965 and his Dr. Tech. Sci. from the Mechanical Engineering Research Institute, Russian Academy of Sciences in 1976. He is currently a Professor and Principal Research Scientist in the Russian Academy of Sciences and a Senior Researcher and Instructor at Naval Postgraduate School in Monterey, California. He is an expert in multicriteria analysis and decision making. He has developed parameter space investigation method (jointly with Dr. Sobol) for correct statement and solution of optimisation problems. This technique has been widely integrated into various fields of industry, science and technology.

Alex Bordetsky received his PhD from Chelyabinsk State Technical University in 1982. He is currently an Associate Professor in the Department of Information Science of Naval Postgraduate School, Monterey, California.

Josef Matusov received his PhD from the Mechanical Engineering Research Institute, Russian Academy of Sciences in 1981. He is currently a Senior Staff Scientist in the Mechanical Engineering Research Institute, Russian Academy of Sciences, Moscow, Russia. He is an expert in uniformly distributed sequences and nets.

Alexander Statnikov received his PhD from Vanderbilt University in 2008. He is currently an Assistant Professor in the Center for Health Informatics and Bioinformatics and Department of Medicine, New York University School of Medicine, New York, New York. He is also the Director of the Computational Causal Discovery laboratory.

1 Introduction

Database search engines are used in essentially all areas of modern life. There are two major types of information retrieval problems that are being solved by search engines. In the first type, a search engine matches a query of the client to a database and reports back results. For example, the client uses PubMed service (http://www.ncbi.nlm.nih.gov/pubmed/) to identify articles that talk about BRCA1 gene in breast cancer using the query 'BRCA1 breast cancer'. Another example is when the client uses Google search engine (http://www.google.com/) to identify a poem from the phrase 'to whose immortal eyes'. Usually this type of problem can be solved efficiently and automatically (i.e., without interaction with the client) given that the database contains a sought solution.

In the second type of problem, the client uses a search engine to retrieve from the database solutions (alternatives) that are described by a set of contradictory characteristics. For example, such problems are widespread when searching for an airline ticket, matching partner, or real estate on the web. In a typical scenario, the database contains a very large number of alternatives described by a set of characteristics. Because of the large number of alternatives, the search engine allows the client to impose constraints on characteristics such that the resulting filtered set of alternatives satisfies all requests of the client. If these constraints are defined correctly, they determine so-called the feasible solution set where the most preferable solution should be then sought by the client. However, in the majority of cases the client defines constraints intuitively and thus often incorrectly. As such, many interesting solutions that under certain assumptions could have been feasible will be lost. This results in the poor or even empty feasible solution set. It is especially the case when the client deals with many contradictory characteristics and many alternatives.[1] Current search engines do not provide tools for systematic construction of the feasible solution set. This is a major shortcoming of existing database search engines, because search for the most preferable solution depends to a large extent on definition of the constraints, and this insurmountable task is shifted to the client's shoulders.

The problem of database search that is addressed in this work involves many contradictory characteristics and a large number of alternatives and necessitates decision-making (Saaty, 2008). The ultimate results also critically depend on the quality of database (Li et al., 2009), type of constraints on characteristics (Stefansen and Borch, 2008) and other factors. An important feature of the database search problem is that filtering of the number of solutions via construction of the feasible solution set should precede decision-making.

That is why the main thesis of this work is that systematic construction of the feasible solution set has a fundamental value for the client. To address this problem, a method called parameter space investigation (PSI) has been created and widely integrated into various fields of industry, science, and technology (Sobol' and Statnikov, 2006; Stadler and Dauer, 1992; Statnikov and Matusov, 1995, 1996, 2002; Statnikov et al., 2008). In this paper, we modify the PSI method to allow construction of the feasible solution set specifically for the problems of database search. We propose a modified PSI method called DBS-PSI (DBS is an acronym for 'database search') that allows to significantly improve the quality of the search results by explicitly constructing and analysing the feasible solution set and identifying the most preferable solutions.

2 Background

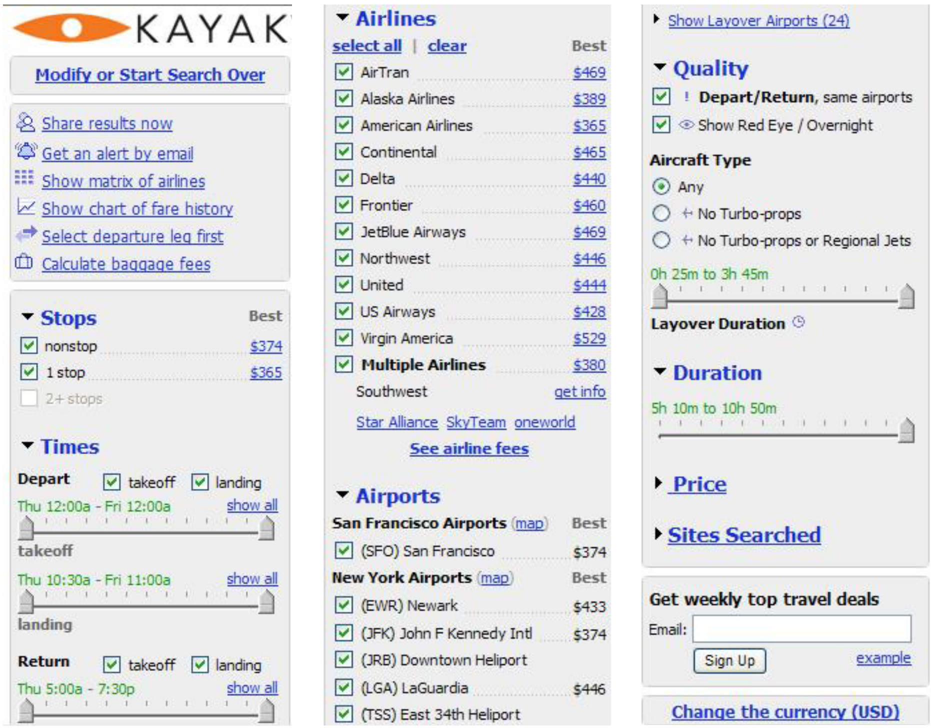
2.1 Characteristics of alternatives: criteria and pseudo-criteria

We define a criterion as the characteristic of the alternative[2] that is related to the alternative's quality by monotonous dependence. In other words, all other things being equal, the more (or the less) the value of the criterion, the better is the alternative. Contrary to the criterion, a pseudo-criterion is not related to alternative's quality by monotonous dependence (Sobol' and Statnikov, 2006; Statnikov and Matusov, 1995; Statnikov et al., 2006; Statnikov et al., 2009). Pseudo-criteria do not need to be optimised; only their constraints have to be satisfied. These constraints are determined either by known and generally accepted standards or by the client's preferences. In most cases, constraints on both criteria and pseudo-criteria are not rigidly set. An algorithm allowing to determine these constraints is provided in the paper. Examples of criteria and pseudo-criteria are given below.

Let us consider four popular uses of database search engines and determine several criteria and pseudo-criteria to illustrate the above definitions:

- Using travel search websites to buy airline tickets, e.g., Kyak travel search engine (http://www.kayak.com). *Criteria*: cost, number of stops, and duration. *Pseudo-criteria*: takeoff time, landing time, airline, airport, see Figure 1. For example, the client most often specifies a range of constraints on takeoff time and landing time. However, if he wanted to depart/arrive as early/late as possible, takeoff time and landing time should be considered as criteria. The choice of destination airport may also be a criterion if the client has to hurry from the airport to work, and as a result he chooses the airport closest to his office.

Figure 1 Kayak travel search engine (see online version for colours)



Stops Best
☑ nonstop $374
☑ 1 stop $365
☐ 2+ stops

Times
Depart ☑ takeoff ☑ landing
Thu 12:00a - Fri 12:00a show all
takeoff
Thu 10:30a - Fri 11:00a show all
landing
Return ☑ takeoff ☑ landing
Thu 5:00a - 7:30p show all

Airlines
select all | clear Best
☑ AirTran $469
☑ Alaska Airlines $389
☑ American Airlines $365
☑ Continental $465
☑ Delta $440
☑ Frontier $460
☑ JetBlue Airways $469
☑ Northwest $446
☑ United $444
☑ US Airways $428
☑ Virgin America $529
☑ **Multiple Airlines** $380
Southwest get info

Airports
San Francisco Airports (map) Best
☑ (SFO) San Francisco $374
New York Airports (map) Best
☑ (EWR) Newark $433
☑ (JFK) John F Kennedy Intl $374
☑ (JRB) Downtown Heliport
☑ (LGA) LaGuardia $446
☑ (TSS) East 34th Heliport

Quality
☑ ! **Depart/Return**, same airports
☑ Show Red Eye / Overnight
Aircraft Type
● Any
○ No Turbo-props
○ No Turbo-props or Regional Jets
0h 25m to 3h 45m
Layover Duration

Duration
5h 10m to 10h 50m

Source: Available at http://www.kayak.com. Fragments

- Using online dating services to look for a matching partner, e.g., Yahoo! Personals (http://personals.yahoo.com/). *Criteria*: education level and income level. *Pseudo-criteria*: appearance and lifestyle. Appearance contains the subsets of more specific pseudo-criteria: ethnicity, height; body type, eye colour, and hair colour. Lifestyle contains the subset of more specific pseudo-criteria: marital status, profession, religion, and interests. In general, when we talk about one's height, weight, and age, it generally makes no sense to say that the higher (lower) are the values of these characteristics the better is the alternative. These characteristics must lie within certain boundaries.

- Using real estate search websites to buy a house, e.g., Realtor.com (http://www.realtor.com/). *Criteria*: cost, distance to work, school district (for school-age children). *Pseudo-criteria*: ceiling height, number of floors. If the client is handicapped and uses a wheel-chair, number of floors becomes a criterion.
- Using online services to buy cars, e.g., Cars.com (http://www.cars.com/). *Criteria*: cost, fuel consumption, operating cost. *Pseudo-criteria*: wear resistance, durability, and operating life of car parts.

2.2 General statement of the problem and solution approach

Assume that a database contains N alternatives (solutions), each of which is described by k characteristics[3] (pseudo-criteria and criteria) Φ_v^j, for $v = 1, \ldots, k$ and $j = 1, \ldots, N$. These characteristics can be contradictory, so improving some characteristics leads to the deterioration of the others. Also assume that the database contains so many alternatives that the client experiences significant difficulties while analysing them and choosing the most preferable alternative. The problem is how to filter the set of alternatives without losing any alternatives that are acceptable to the client. This can be accomplished by constructing the feasible and Pareto optimal sets.

We are given a vector of characteristics $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_k)$, where $\Phi_1, \Phi_2, \ldots, \Phi_c$ are criteria that we want to minimise and $\Phi_{c+1}, \Phi_{c+2}, \ldots, \Phi_k$ are pseudo-criteria. Alternatives Φ^j that:

1 satisfy all criteria constraints, i.e., $\Phi_v^j (v = 1, \ldots, c)$ such that $\Phi_v^j \le \Phi_v^{**}$ (or $\Phi_v^j \ge \Phi_v^{**}$ in case of maximisation)

2 satisfy constraints on pseudo-criteria, i.e., $\Phi_v^j (v = c+1, \ldots, k)$ such that $\Phi_v^* \le \Phi_v^j \le \Phi_v^{**}$,

constitute the feasible solution set D.

Here Φ_v^{**} is the worst value of criteria which the client considers as acceptable; constraints on pseudo-criteria Φ_v^* and Φ_v^{**} correspond to the best and worst values, respectively.

We solve the general problem by first constructing the feasible solution set D by imposing constraints on criteria and pseudo-criteria mentioned above via dialogs of the client with computer. Next, we define Pareto optimal set $P \subseteq D$ using only criteria Φ_1, $\Phi_2, \ldots, \Phi_c$. Recall that an alternative $\Phi_v^{j_0} \in D$ is called Pareto optimal if there exists no alternative $\Phi^j \in D$ such that $\Phi_v^j \le \Phi_v^{j_0}$ for all $v = 1, \ldots, c$ and $\Phi_{Nv_0}^j < \Phi_{Nv_0}^{j_0}$ for at least one $v_0 \in \{1, \ldots, c\}$. In other words, alternative Φ^{j_0} cannot be improved by all criteria $\Phi_1, \Phi_2, \ldots, \Phi_c$ simultaneously. A set $P \subseteq D$ is called a Pareto optimal set if it consists of only Pareto optimal alternatives. Finally, the client determines the alternative $\Phi_v^{j_0} \in P$ which is the most preferable among the alternatives belonging to the set P (Lichtenstein and Slovic, 2006).

2.3 Motivation of the problem statement

Concessions Φ_v^* and Φ_v^{**} are something like a market where we have the opportunity to trade until we get what we need. Of course, not all concessions are possible, but that is outside the scope of our work. By investigating the set of Pareto solutions, the clients see what can and what cannot be achieved. They are able to choose the most preferable alternative in this set, $\Phi^{j_0} \in P.$ At the same time, they know that there exist no better alternatives in the database! The Pareto optimal set plays an important role in vector optimisation problems because it can be analysed more easily than the feasible solution set and because the optimal alternative always belongs to the Pareto optimal set, irrespective of the system of preferences used by the client for comparing alternatives belonging to the feasible solution set.

2.4 Problem considered by this work

This paper primarily considers problems with the following features:

1. characteristics are contradictory
2. the dimensionality of the characteristics vector may reach dozens
3. the number of alternatives may reach many thousands
4. the constraints on these characteristics are often stringent
5. the constraints on characteristics Φ_v^* and Φ_v^{**} can be defined in an interactive mode, in the process of dialogs of the client with computer.

2.5 What do the client and search engine developers need to know?

Search engine developers must provide all necessary characteristics (criteria and pseudo-criteria). The client must understand which constraints on characteristics are acceptable and which are not, without differentiating between criteria and pseudo-criteria when constructing the feasible set. After determining the feasible set (feasible alternatives), the client himself must decide which of these characteristics is a criterion and which one is a pseudo-criterion. This is because the same characteristics may be a criterion for one client and pseudo-criteria for another client. Next comes the analysis of the obtained alternatives and selection of the most preferable one.

3 Methods

3.1 DBS-PSI method as a new paradigm of database search

We mentioned earlier difficulties in constructing the feasible solution set correctly. To do this, the client must see what concessions he should make on characteristics and what he will get in return.[4] The purpose of our work is to show the process of generating concessions to the client and to interactively engage him in it.

Recall that the PSI method was designed to help the client determine constraints on characteristics (criteria and pseudo-criteria) and thus to construct the feasible solution set. After determining the feasible solution set, the Pareto optimal set is constructed only with consideration of the criteria. Then the client will search for the most preferable solution on the Pareto optimal set (Sobol' and Statnikov, 2006; Statnikov, 1999; Statnikov and Matusov, 2002). The new DBS-PSI method is designed to be used for database search, and it consists of three main stages and are presented below.

Stage 1 Compilation of analysis tables via computer

In this stage, the client transforms the database into analysis tables[5] as follows. For each characteristics an analysis table is compiled so that the values of $\Phi_v^1,\ldots,\Phi_v^N$ are arranged in increasing order; i.e.,

$$\Phi_v^{i_1} \le \Phi_v^{i_2} \le \ldots \le \Phi_v^{i_N}, v = 1,\ldots k \tag{1}$$

where i_1, i_2, …, i_N are the numbers of alternatives (a separate set for each v). Taken together, the k tables form complete analysis tables. In analysis tables, the list of non-numerical criteria is arranged in order of the client's preferences.

Stage 2 Selection of characteristic constraints

This stage includes the dialogue of a client with the computer. Let $\min\Phi_v$ and $\max\Phi_v$ be *unfeasible solutions* for the v-th characteristic (pseudo-criterion and/or criterion), and let Φ_v^* and/or Φ_v^{**}, respectively, be the *minimum possible concessions* from $\min\Phi_v$ and/or $\max\Phi_v$, $(\min\Phi_v \le \Phi_v^* \le \Phi_v^{**} \le \max\Phi_v)$, where $(\min\Phi_v;\Phi_v^*]$ and $(\Phi_v^{**};\max\Phi_v]$ are *unfeasible solution* domains. Then the remainder of table $[\Phi_v^*;\Phi_v^{**}]$ is the *maximum* search interval for feasible solutions. After examining the characteristics arranged in each of the analysis tables in increasing (decreasing) order, starting with $\min\Phi_v$ and/or $\max\Phi_v$, the client analyses each value. The first feasible value found will be $\Phi_v^*(\Phi_v^{**})$. If $\min\Phi_v$ and/or $\max\Phi_v$ are feasible values, there is no need to make the corresponding concessions. In general, we will denote the feasible solution search interval as $[\Phi_v^*;\Phi_v^{**}]$. If the selected values of $[\Phi_v^*;\Phi_v^{**}]$ are not a maximum, then many interesting solutions may be lost, since some of the characteristics are contradictory.

The client has to consider one characteristic (analysis table) at a time and specify the respective constraints Φ_v^* and Φ_v^{**}. Then, he proceeds to the next analysis table, and so on. Note that revision of the characteristic constraints does not cause any difficulties to the client.

It is important to emphasise that a human-computer dialog is very convenient for the client: he does not have to change the values of some characteristics at the expense of others. He sees one analysis table and sets the appropriate constraints; then he repeats the same process with another table, and so on. After the client defines the constraints, the computer searches for the feasible solutions. For example, if the client has defined the constraints on the first and second characteristics, he *immediately* obtains information on feasible solutions based on these characteristics. After defining the constraints on the third characteristic, he obtains feasible solutions based on three characteristics, and so on. In other words, the client sees what *price he pays for making concessions on various characteristics, i.e., what he loses and what he gains*. Such an analysis gives the client valuable information on the advisability of revising various characteristic constraints with the aim of improving the basic characteristics. Thus, in this method, the client imposes constraints on each characteristic in succession and follows the construction of the feasible solution set step-by-step. Moreover, he can go back at any time and re-examine the constraints depending on the feasible solution set obtained.

Stage 3 Verification of the solvability of a problem via computer

Let the client fix a characteristic, e.g., Φ_{v_1}, and consider the corresponding analysis table (1). Let S_1 be the number of values in the table satisfying the selected characteristic constraints:

$$\Phi_{v_1}^{*} \le \Phi_{v_1}^{i_1} \le \ldots \le \Phi_{v_1}^{i_{s1}} \le \Phi_{v_1}^{**} \tag{2}$$

Then characteristic Φ_{v_2} is selected by analogy with Φ_{v_1} and the values of $\Phi_{v_2}^{i_1},\ldots,\Phi_{v_2}^{i_{s2}}$ in the analysis table are considered. Let the table contain $S_2 \le S_1$ values such that $\Phi_{v_2}^{*} \le \Phi_{v_2}^{i_j} \le \Phi_{v_2}^{**}$, where $1 \le j \le S_2$. Similar procedures are carried out for each characteristic. If at least one alternative can be found for which all characteristic constraints are valid simultaneously, then the feasible solution set is non-empty and our problem is solvable. Otherwise, the client should return to Stage 2 and make certain concessions on characteristic constraints Φ_{v}^{*} and Φ_{v}^{**}. The procedure is iterated until the feasible solution set is non-empty. Now the Pareto optimal set is constructed. This is done by removing those feasible alternatives that can be improved with respect to all characteristics simultaneously.

As a rule, the client makes more concessions on the less important characteristics. These concessions may allow him to obtain a substantial gain in the most significant characteristics.

Notice that the number of characteristics must be no less than necessary. The client always wishes to optimise not one, but all of the most important characteristics, many of which are antagonistic. The greater the number of characteristics is taken into account, the richer is the information obtained about alternatives. The DBS-PSI method allows us to consider as many characteristics as necessary.

4 Example: searching for a matching partner

The purpose of this example is to give readers an idea of how to construct feasible and Pareto optimal solutions on the basis of analysis tables. Earlier we mentioned the example of using online dating services to look for a matching partner. Let us show how the client will search for a matching partner (woman) using the DBS-PSI method. Suppose the search characteristics are height (Φ_1), weight (Φ_2), age (Φ_3), income (Φ_4) and education (Φ_5). Let the client have access to a database with alternatives. According to Stage 1 of the DBS-PSI method, the client should construct analysis tables, see Figure 2. From Figure 2, it can be seen that the database contains a large number of alternatives[6] and the values of the characteristics vary over fairly large ranges: from a height of 4’11” (alternative #284) to 6’11” (alternative #76), from a weight of 99.2 lbs (alternative #5298) to 183 lbs (alternative #66), from 25 years of age (alternative #971) to 65 years of age (alternative #34006), from an income of $25,000 (alternative #1021) to $150,000 (alternative #90661), and from an education of some high school (alternative #12907) to postgraduate (alternative #95175). The more alternatives there are in the analysis tables, the greater is the client’s choice!

Figure 2 Fragment of analysis tables (see online version for colours)

#	*Height, Φ_1*	#	*Weight, Φ_2*	#	*Age, Φ_3*	#	*Income, Φ_4*	#	*Education, Φ_5*
284	4’11”	5,298	99.2 lbs	971	25 years	1,021	$25,000	12,907	Some high school
…	…	…	…	…	…	…	…	…	…
689	5’2”	88	125.7 lbs	472	30	894	$25,000	9,705	High school graduate
…	…	…	…	…	…	…	…	…	…
27	**5’3”**	…	…	…	…	…	…	12,096	High school graduate
63	5’3”	**7,632**	**127.9 lbs**	…	…	…	…	63	Some college
…	…	97	130.1 lbs	…	…	14,007	$35,000	…	…
12,089	5’3”	…	…	888	34	7	$50,000	…	…
92	5’4”	894	130.1 lbs	**7,632**	**34**	…	…	…	…
…	…	…	…	…	…	**27**	**$50,000**	**7,632**	**Some college**
894	5’4”	10	132.3 lbs	9	35	80,092	$50,000	**27**	**College graduate**
…	…	…	…	…	…	21	$75,000	…	…
1,964	5’4”	1,093	138.9 lbs	23	36	…	…	…	…
38	5’5”	…	…	894	36	…	…	…	…
7,632	**5’5”**	…	…	…	…	…	…	…	…
30,617	5’5”	**27**	**143.3 lbs**	…	…	**7,632**	**$75,000**	…	…
…	…	…	…	**27**	**38**	245	$100,000	…	…
994	5’6”	…	…	…	…	…	…	…	…
…	…	30,617	150 lbs	300	40	6,005	$100,000	88,945	College graduate
…	…	…	…	…	…	711	$150,000	992	Postgraduate
…	…	…	…	…	…	…	…	…	…
76	6’11”	66	183 lbs	34,006	65	90,661	$150,000	95,175	Postgraduate

Note: To simplify visualisation, values of characteristics are replaced with ‘…’ in many places.

According to Stage 1, the client should construct search intervals for feasible solutions. After studying the analysis table for the first characteristic (height), the client has determined the search interval for feasible solutions $[\Phi_1^*;\Phi_1^{**}]=[5’2”;5’6”]$, see Figure 2. Alternatives #689,…,#27,…,#994 correspond to this interval. The client then proceeds to the second analysis table, to weight, and determines the search interval for feasible solutions $[\Phi_1^*;\Phi_1^{**}]=[125.7\text{ lbs};150\text{ lbs}]$. Alternatives #88,…, #7632, …, #30617 correspond to this interval. When switching from one analysis table to another (from height to weight), the computer records the total number of alternatives that satisfy the constraints on these characteristics simultaneously. As it turns out, there are *four* of these alternatives, #27, #894, #7632 and #30167. If there were no feasible alternatives, client should revise the constraints on the first and/or second characteristics. The client then proceeds to the third analysis table (age) and constructs the interval $[\Phi_3^*;\Phi_3^{**}]=[30\text{ years};40\text{ years}]$ that is satisfied by alternatives #472, …, #888, …, #300. The total number of alternatives that satisfies the constraints on three characteristics simultaneously is reduced to three, #27, #894 and #7632. Then the client examines income (the fourth analysis table). Not a single alternative has satisfied the constraints $150,000, and $100,000. After setting constraint to $75,000, the alternative #7632 is found that satisfies all constraints on four characteristics. Finally, the client reduces the constraint to $50,000. As a result, the total number of alternatives that satisfy constraints on four characteristics is *two*, namely, #27 and #7632, see Figure 2.

In the last analysis table (the ‘education’ characteristic), not a single alternative satisfied the ‘postgraduate’ constraint. After setting the constraint to ‘some college’, *two feasible alternatives* (*matching partners*) #27 and #7632 are obtained (shown with bold in the Figure 2). These alternatives satisfy all constraints on five characteristics. The values of characteristics of matching partners #27 and #7632 also are given in Table 1. All constraints and matching partners were determined in an interactive mode.

In Figure 2, the areas of search for feasible alternatives are highlighted in tan (pseudo-criteria) and in light green (criteria) accordingly.

Table 1 Input data and some search results

Characteristics	*Height Φ_1*	*Weight, lbs, Φ_2*	*Age, years, Φ_3*	*Income, $, Φ_4*	*Education Φ_5*
Range of characteristic values	4’11”–6’11”	99.2–183	25–65	25,000–150,000	Some high school–Postgraduate
Intervals of search for feasible solutions	5’2”–5’6”	125.7–150	30–40	50,000–150,000	Some college–Postgraduate
Feasible alternatives (matching partners)					
#27	5’3”	143.3	38	50,000	College graduate
#7632	5’5”	127.9	34	75,000	Some college

4.1 Conclusions of the example

Constructing and studying analysis tables allowed to obtain two feasible alternatives. After that the client may wish to determine the optimal alternative set with consideration of all or the most important criteria. For example, if the client had searched for the optimal solutions according to the criterion income, the search engine would have given him alternative #7632; alternative #27 would be better for the criterion education. If two criteria, income and education, are considered simultaneously, the Pareto optimal solution set will contain the two specified solutions.

However, according to Yahoo! Personals (http://personals.yahoo.com/), upon starting a search, the client should impose constraints on age and geographical area, see Figure 3.

In our opinion, these recommendations are incorrect, since the client can lose compromise solutions that are acceptable to him. These constraints should be defined in the process of studying the analysis tables; after finding the feasible alternatives (matching partners) in an interactive mode, the remaining solutions will automatically become unfeasible.

Figure 3 Yahoo! Personals search engine (see online version for colours)

I'm a Man seeking a Woman Age 30 to 40

within 50 miles of Boston, MA [change]

Notes: available at http://personals.yahoo.com/. Fragment.

5 Conclusions

It is often the case that database search involves many contradictory criteria and many alternatives, and it is necessary to search for compromise solutions. This can be efficiently accomplished by construction and analysis of the feasible solution set. However, current search engines do not systematically construct the feasible solution set, thus 'hiding' many potentially interesting solutions from the client. To address this problem, we introduced the DBS-PSI method which allows the client to:

a construct analysis tables

b determine the feasible solution set in the process of dialogs with the computer

c find Pareto optimal set

d identify the most preferable solution.

The process of searching for the feasible solution set based on analysis tables is client-friendly: he sees what he loses and what he gains for every concession. The DBS-PSI method as a new paradigm of database search can be used in all areas of modern life and promises to increase efficiency of modern search engines. The relevancy of the proposed methodology is apparent given ongoing expansion of the World Wide Web and search engines.

Acknowledgements

The authors are grateful to Professor B. Peltsverger (Georgia Southwestern State University, USA), Professor Peter Ekel (University of Minas Gerais, Brazil), Dr. Terry E. Smith (Naval Postgraduate School, USA), Karen Olson, and Irina Statnikova for interest in our work.

References

Li, X., Zhang, L., Zhang, P. and Shi, Y. (2009) 'Problems and systematic solutions in data quality', *International Journal of Services Sciences*, Vol. 2, No. 1 pp.53–69.

Lichtenstein, S. and Slovic, P. (2006) *The Construction of Preference*, Cambridge University Press.

Saaty, T. (2008) 'Decision making with the analytic hierarchy process', *International Journal of Services Science*, Vol. 1, No. 1 pp.83–98.

Sobol', I.M. and Statnikov, R.B. (2006) *Selecting Optimal Parameters in Multicriteria Problems*, 2nd ed., in Russian, Drofa, Moscow.

Stadler, W. and Dauer, J.P. (1992) 'Multicriteria optimization in engineering: a tutorial and survey. Structural optimization: status and promise', in Kamat, M.P. (Ed.): *Progress in Astronautics and Aeronautics, American Institute of Aeronautics and Astronautics*, Vol. 150, pp.209–249, Washington, DC.

Statnikov, R.B. (1999) *Multicriteria Design. Optimization and Identification*, Kluwer Academic Publishers.

Statnikov, R.B. and Matusov, J.B. (1995) *Multicriteria Optimization and Engineering*, Chapman & Hall, New York.

Statnikov, R.B. and Matusov, J. (1996) 'Use of P_τ nets for the approximation of the Edgeworth-Pareto set in multicriteria optimization', *Journal of Optimization Theory and Applications*, Vol. 91, No. 3, pp.543–560.

Statnikov, R.B. and Matusov, J.B. (2002) *Multicriteria Analysis in Engineering*, Kluwer Academic Publishers.

Statnikov, R.B., Ali, A.K., Bordetsky, A. and Statnikov A. (2008) 'Visualization approaches for the prototype improvement problem', *Journal of Multi-Criteria Decision Analysis*, Vol. 15, pp.45–61.

Statnikov, R.B., Bordetsky, A. and Statnikov, A. (2006) 'Multicriteria analysis tools in real-life problems', *Journal of Computers & Mathematics with Applications*, Vol. 52, Nos. 1–2, pp.1–32.

Statnikov, R.B., Bordetsky, A. and Statnikov, A. (2009) 'Management of constraints in optimization problems', *Journal of Nonlinear Analysis*, Vol. 71, No. 12, pp.e967–e971.

Stefansen, C. and Borch, S.E. (2008) 'Using soft constraints to guide users in flexible business process management systems', *International Journal of Business Process Integration and Management*, Vol. 3, No. 1 pp.26–35.

Notes

1 In rare cases, when there are a few alternatives and characteristics, this task may be realistic for the client.

2 Examples of alternatives in database search engines are airline tickets (travel search websites), women/men (online dating services), houses (real estate search websites), cars (online services to buy cars), and so on.

3 It is worth noting that often client can a priori state 'rigid' requirements to certain characteristics that cannot be changed under any circumstances. In such cases, when searching for the most preferable solution, it is necessary to exclude alternatives that did not satisfy those requirements and then search for feasible solutions as described in the paper.

4 We note that in search engines it is possible to revise constraints. However, the client is expected to do this at very intuitive level.

5 The foundation of the PSI method is so-called test tables. The principal difference between test tables and analysis tables is that in the former case, solutions are obtained with the use of generators (e.g., uniformly distributed sequences), and in the latter case from a database. Also, the parameters are absent in the search engines that are discussed in this paper.

6 For example #12089, #34006, #95175, #30617, #88945 and so on. Each of the women in the analysis tables is designated as alternative #.