A Classification of Mutational Approaches for Genetic Search

Drazen Draskovic, Bosko Nikolic, Veljko Milutinovic
Faculty of Electrical Engineering, University of Belgrade
drazen.draskovic@etf.bg.ac.rs, bosko.nikolic@etf.bg.ac.rs, veljko.milutinovic@etf.bg.ac.rs

Abstract - Genetic algorithms for Internet Search were classified a lot in the open literature, but one specific aspect there off - the mutational approaches - was not. This paper represents an effort to shot light on the existing mutational approaches in the context of the genetic algorithms that they are a part of. Major contributions of this paper are: (a) An original classification, which opens some potentially fruitful research avenues; (b) A block diagram based representation of the four major classes of the newly introduced classification; (c) Uniform pseudo-code based presentation of selected algorithms, in a way that enables easy comparison; (d) Discussion of essential issues in a way that opens up new avenues for future research.

I. INTRODUCTION

Genetic search algorithms have been classified and surveyed in a number of different papers; however, there is no extensive effort in the open literature to classify and survey mutational approaches for genetic search. Even more, there is no effort to classify them in a way that opens up new research avenues in the domains of interdisciplinary and multidisciplinary domains, with a stress on applications in computer monitoring and complex control.

In this work, under the term multidisciplinary, we imply the hybrid efforts that combine characteristics of two or more approaches. Under the term interdisciplinary, we imply symbiotic efforts that take the best of two or more approaches, and generate new solutions somewhere in the space in-between.

The starting point of this paper and the related research is the coursework at the University of Belgrade, in two different areas: Artificial Intelligence and Knowledge Search. Among the major goals of these two courses are the efforts to teach students how to think and create in domains of interdisciplinary and multidisciplinary, with special emphasis on monitoring and control.

The rest of this paper is organized as follows: Section 2 considers two basic classification criteria. Section 3 describes the proposed classification. Section 4 presents basic class characteristics, and Section 5 outlines directions for future work. Section 6 concludes the paper.

II. CLASSIFICATION CRITERIA

For the benefit of this effort, two basic classification criteria were selected: (a) Using mutational databases, and (b) Doing concept oriented analysis. These criteria were selected both because they reflect some recent and successful research efforts, and also because they leave space for interdisciplinary and multidisciplinary approaches to future research.

The secondary classification criteria were chosen to be different for two different branches of the basic classification: (a) In the database branch, mutation can be based on a random approach or on a targeted approach, and (b) In the analysis branch of the classification, mutation can be based on semantic analysis or on data-mining analysis.

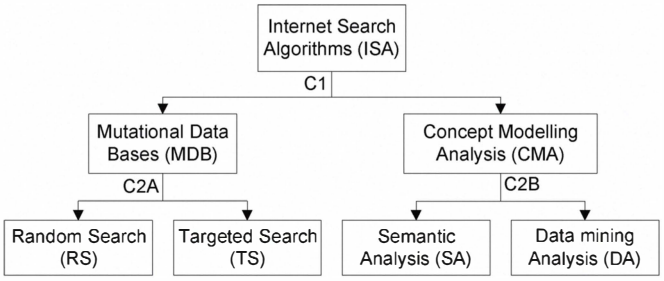

Figure 1. Classification of Internet Search Algorithms

Legend: C1 (criterion #1) = retrieval-oriented vs analysis-oriented
C2A (criterion #2, in the MDB path) = Random Search vs Targeted Search
C2B (criterion #2, in the CMA path) = Semantics-oriented vs Datamining-oriented
Note: Subclasses (based on further criteria) are also possible.

III. PROPOSED CLASSIFICATION

Figure 1 presents the proposed classification and defines the terms of interest. Further subclasses are also possible, in each one of the four major classes. Elaboration of further subclasses is a subject of a follow-up research.

The proposed classification enables approaches to be combined, to create new research avenues. In the interdisciplinary domain, if only two approaches are combined, there are 6 different possibilities. The same number of new research possibilities can be created in the multidisciplinary method that combines two approaches. The most promising ones are discussed in Section 5. A summary is given in Table 1.

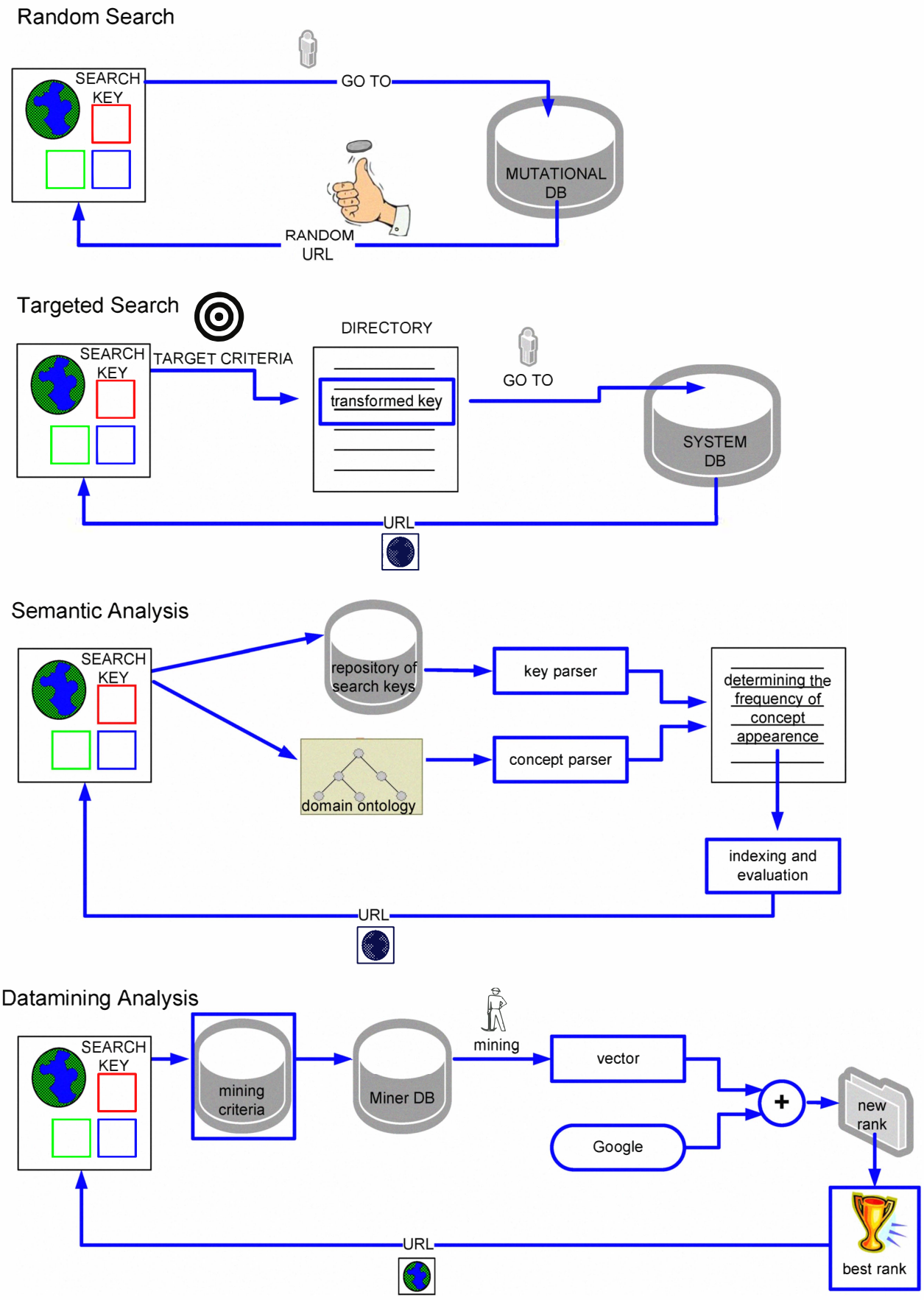

Figure 2. Generalized Structure of the four Search Classes

Legend: DB = Database;
URL = type of URI that's used to describe the location of a specific document;
Note: All used symbols were chosen to be intuitive. Essential explanations are in the basic text of this paper.

Table 1 defines technical and symbolic names of all major classes, and lists the starting-point literature, followed by a set of selected references belonging to each one of the four major classes. Elaboration is a subject of a follow-up paper.

IV. BASIC CLASS CHARACTERISTICS

Figure 2 represents block diagrams of the generalized structures of the four major classes. Both the data flow and computation flow are described in an intuitive way, for easy understanding of the notions behind.

Random search using a mutational database was used in some early research in the field, with stress on scientific environments (see Figure 3).

```
RANDOM SEARCH(key)
find_key_in_mutational_db(key)
mutation()
return random_url
```

Figure 3. Pseudo-code of the Generalized Procedure of the Algorithms utilizing the Random Search approach

Targeted research with locality oriented mutation on system databases was used in some mission critical business oriented environments (see Figure 4).

```
TARGETED SEARCH(key)
find_key_in_directory()
mutation()
find_transformed_key_in_db()
return url
```

Figure 4. Pseudo-code of the Generalized Procedure of the Algorithms utilizing the Targeted Search approach

Mutations based on semantic analysis can be used in social networks and professional for searching, as well as in the engineering aspects of monitoring and control (see Figure 5).

Mutations that data-mine the satisfaction and profile of customers can be relatively successful in commercial environments or in educational applications (see Figure 6).

The presented pseudo-codes show the essence of the utilized mutation types, in the context of entire genetic search algorithms. The instances at which mutations take placed are italized.

```
SEMANTIC SEARCH(key)
i, j = 0
find_in_repository_of_search_keys(key)
read_ontology_and_find_synonims()
format_key_list()
m = size_key_list()
if (i < m) then
   format_concept_list()
   n = size_concept_list()
   if (j < n ) then
      read_concept()
      mutation()
      determining_frequency()
      indexing_and_evaluation()
      j = j + 1
   end_if
   i = i + 1
end_if
return url
```

Figure 5. Pseudo-code of the Generalized Procedure of the Algorithms utilizing the Semantic Analysis approach

```
DATAMINING SEARCH(key)
i = 0
find_mining_criteria()
find_key_in_miner_db(key)
format_vector()
n = vector_size()
if (i < n)
   crossover_vector_and_criteria()
   mutation()
   insert_in_list_new_rank_vector()
   i = i + 1
end_if
find_best_rank_in_new_list()
return url
```

Figure 6. Pseudo-code of the Generalized Procedure of the Algorithms utilizing the Datamining Analysis approach

V. OPEN AVENUES FOR FUTURE RESEARCH

When it comes to future research of the interdisciplinary/multidisciplinary nature, in the mutational data base branch, there is a set of conditions when random search works better, and another set of conditions when targeted research works better. If we add targeting oriented mechanisms into a mutational database, we obtain an interdisciplinary approach. If we add a mechanism that tests the conditions of the application environment, and another one that, based on the conditions, selects either the random or the targeted approach, we obtain a multidisciplinary solution.

In the first case we have a symbiotic approach; in the second case we have a hybrid approach.

Similarly, in the concept modeling analysis branch, in some/other conditions, one/other approach works better. Analogously, in these conditions, both a symbiotic and a hybrid solution can be devised, along the lines of general interdisciplinary and multidisciplinary research methodologies.

The above text describes two possible symbiotic and two possible hybrid approaches. As indicated above, the plethora of possible approaches is much wider. Elaboration of the remaining four approaches in each one of the two branches is a subject of a follow up paper.

All the approaches mentioned above can be compared analytically or by simulations, to see what benefits they bring. The benefits in the performance domain are obvious (combining the best of two different approaches can bring only benefits). However, combining two approaches inevitably brings a complexity increase (both approaches plus selection/combining mechanisms require more transistor count than one isolated approaches). Consequently, the crucial issue here is the price/performance analysis.

Technical Names	Random Search (RS or RS/MDB)	Targeted Search (TS or TS/MDB)	Semantic Analysis (SA or SA/CMA)	Data-mining Analysis (DA or DA/CMA)
Symbolic Names	Lion	Jaguar	Tiger	Panthera
Number of Surveyed Contributions	2	4	4	4
References	[Nikolic2011a] Nikolic, B., "Expert Systems", WUS Austria Educational Publishing and University of Belgrade, Classroom Textbook, June 2011.	[Milutinovic2000a] Milutinovic, V., Cvetkovic, D., Mirkovic, J., "Genetic Search Based on Multiple Mutation Approaches", IEEE Computer, November 2000, vol. 33, issue: 9, pp. 118-119	[Nikolic2011b] Furlan, B., Sivacki, V., Jovanovic, D., Nikolic, B., "Comparable Evaluation of Contemporary Corpus-Based and Knowledge-Bases Semantic Similarity Measures of Short Text", JITA, vol. 1, no. 1, pp. 65-72, ISSN: 2232-962, June 2011.	[Milutinovic2000b] Milutinovic, V., Knezevic, P., Radunovic, B., Casselman, S., Schewel, J., Obelix Searches Internet Using Customer Data, IEEE Computer, July 2000, vol. 33, issue: 7, pp. 104-107
	[Nick2001]	[Simon2009] [Mirkovic1999] [Chen1997]	[Gordon2006] [Leroy2003] [Wang2006]	[Al-Dallal2009] [Hu2007] [Freitas2001]

TABLE I: SUMMARY OF THE APPROACHES THAT LED TO THE CLASSIFICATION PROPOSED IN THIS PAPER

Legend: Technical names are obtained by combining the names and abbreviations of the applied classification criteria. Symbolic names have been chosen with characteristics that associate to those of the related classes.

Note: Hybrid approaches are also possible, of both the multidisciplinary type (including characreristics of two or more classes) and the intradisciplinary type (with characteristics between those of two or more classes). Hybrid approaches open new avenues for future research.

VI. CONCLUSION

In this research, we have presented a detailed classification of algorithms for Internet search, using two basic classification criteria. The proposed classification allows the possibility of combining the defined classes of algorithms and creates opportunities for new "hybrid" algorithms. Also, by combining the proposed classes of algorithms, the following research can get higher performance of algorithms for Internet search.

In addition, if we exclude the use of these algorithms in search engines, which represents their main purpose, the benefit of these algorithms is in the student education domain, because the implementation of these algorithms can be very important for educational purposes, in the courses on artificial intelligence. As recommended by expert organizations IEEE Computer Society and ACM, starting in the year 2005, the area of genetic algorithms is proposed on the list of mandatory topics in artificial intelligence courses at the undergraduate or master studies, for students of Computes Science [IEEE2005].

The newly open problems are: (a) To elaborate, by adding a pseudo-code for each presented class; (b) To develop a tool that enables all presented examples to be compared; (c) To develop a tool that enables education of students on GA in general and MA in specific and (d) which is the main contribution of this paper, to develop original examples along all symbiotic and synergistic combinations that make sense.

ACKNOWLEDGMENT

This work is supported by the Ministry of Science, Republic of Serbia, grant TR32047.

REFERENCES

[Al-Dallal2009] Al-Dallal, A.; Shaker, R., "Genetic Algorithm in Web Search Using Inverted Index Representation", 5th IEEE GCC Conference & Exhibition, March 17-19, 2009, Kuwait City, on pages 1-5, Print ISBN: 978-1-4244-3885-3

[Chen1997] Chen, H.; Chung, Y.; Ramsey, M.; Yang, C.; Ma, P.; Yen, J., "Intelligent Spider for Internet Searching," Proceedings of the Thirtieth Annual Hawaii International Conference on System Sciences, Maui, Hawaii, USA, January 1997.

[Freitas2001] Freitas, A., A Survey of Evolutionary Algorithms for Datamining and knowledge discovery, Advances in Evolutionary Computation, conference, 2001, pages 819-845

[Gordon2006] Gordon, M.; Fan, W.; Pathak, P., Adaptive Web Search: Evolving a Program That Finds Information, IEEE Intelligent Systems, September/October 2006, vol. 21, no. 5, pp. 72-77

[Hu2007] Hu, M.; Lim, E.; Sun, A.; Lauw, H.W.; Vuong, B., "On Improving Wikipedia Search using Article Quality", WIDM '07 ACM International Workshop on Web information and data management, November 9, 2007, Lisboa, Portugal, on pages 145-152, Print ISBN: 978-1-59593-829-9

[IEEE2005] http://www.acm.org/education/education/curric_vols/CC2005-March06Final.pdf, accessed on September 8th, 2010

[Leroy2003] Leroy, G.; Lally, A.M.; Chen H., "The Use of Dynamic Contexts to Improve Casual Internet Searching", ACM Transaction on Information Systems, July 2003, vol. 21, no. 3, pages 229-253

[Milutinovic2000a] Milutinovic, V.; Cvetkovic, D.; Mirkovic, J., "Genetic Search Based on Multiple Mutation Approaches", IEEE Computer, November 2000, vol. 33, issue: 9, pp. 118-119
IEEE Computer, July 2000, vol. 33, issue: 7, pp. 104-107

[Milutinovic2000b] Milutinovic, V.; Knezevic, P.; Radunovic, B.; Casselman, S.; Schewel, J., Obelix Searches Internet Using Customer Data, IEEE Computer, July 2000, vol. 33, issue: 7, pp. 104-107

[Mirkovic1999] Mirković, J.; Cvetković, D.; Nešić, Lj.; Tomča, N.; Cvetićanin, S.; Slijepčević, S.; Obradović, V.; Mrkić, M.; Čakulev, I.; Milutinović, V.; Kraus, L., "Genetic Algorithms for Intelligent Internet Search: A Survey and a Package for Experimenting with Various Locality Types," IEEE TCCA Newsletter, 1999.

[Nick2001] Nick, Z.Z.; Themis, P., "Web Search using a Genetic Algorithm", Internet Computing IEEE, 2001, vol. 5, issue: 2, pp. 18-26

[Nikolic2011a] Nikolic, B., "Expert Systems", WUS Austria Educational Publishing and University of Belgrade, Classroom Textbook, June 2011.

[Nikolic2011b] Furlan, B.; Sivacki, V.; Jovanovic, D.; Nikolic, B., "Comparable Evaluation of Contemporary Corpus-Based and Knowledge-Bases Semantic Similarity Measures of Short Text", JITA, June 2011, vol. 1, no. 1, pp. 65-72, ISSN: 2232-962

[Simon2009] Simon, P.; Sathya, S.S., "Genetic algorithm for information retrieval", Intelligent Agent & Multi-Agent Systems, International Conference, Chennai, July 2009, pp. 1-6

[Wang2006] Wang, J.; Ding, Z.; Jiang, C., "GAOM: Genetic Algorithm based Ontology Matching", IEEE Asia-Pacific Conference on Services Computing (APSCC '06), Guangzhou, Guangdong, December 2006, pages 617-620, Pring ISBN: 0-7695-2751-5