International Journal of Computational Intelligence and Informatics, Vol. 5: No. 2, September 2015

Improving the Navigability of Web Pages using Genetic Algorithm

Mrs. N. Thangammal
Teaching Assistant,
Dept. of Computer Science,
Madurai Kamaraj University College
Madurai, Tamil Nadu, India

Dr. A.Pethalakshmi
Associate Professor & Head,
Dept. of Computer Science
M.V.M.Govt. Arts College for Women
Dindigul, Tamil Nadu, India

***Abstract* - With the recent spurt in the number of engineering web applications, the challenge of improving the quality of the engineered applications has come to the forefront. One important aspect of the quality of a web application is navigability. It represents the ease and convenience with which the user of the web site can have quick access to all the information contained in the web site. Genetic algorithms that mimic the natural process of evolution to uncover solutions to problems have achieved huge success with complex problems across multiple domains. This paper presents a genetic algorithmic approach to improving the navigability of web pages that identifies the best sequence of transformations that can improve navigability.**

Keywords – Navigability, Genetic Algorithm, Web Engineering, Quality Model

I. INTRODUCTION

The past decade has witnessed an avalanche in the number of the engineered web applications. Given that web applications provide a business advantage to organizations, the quality of the applications is of paramount importance [12]. Business organizations prefer to be equipped with web sites that provide a pleasant and convenient browsing experience to users.

The Quality challenges imposed by web applications are different from those posed by ordinary applications [9][10]. While quality of web applications encompasses a large number of aspects, navigability is perceived to be one of the most important. Navigability refers to the relative ease with which a user can have access to all the information he needs from the web site. The present paper proposes an approach to improve navigability by identifying the best sequence of transformations that can be applied to a web page to improve its navigability.

II. RELATED WORK

Over the years, many researchers have sought to improve navigability of web pages and these have achieved reasonable success [9[12][8]. Data Mining [11] can be used for the purpose. An approach for addressing web usability issues has been proposed by Nielsen and Loranger [1]. A set of metrics that can be used to assess Web Application quality and evolvability has been proposed by Boldyreff et.al. [2]. Olsina et al. propose WebQEM (Web Quality Evaluation Methodology) [3] and Albuquerque et al. propose FMSQE (Fuzzy Model for Software Quality Estimation) [4]. The work of Vaucher, Boclinville, Sahraoui and Habra on recommending improvements to web applications using simulated annealing [5] is directly related to the present work.

III. THE QUALITY MODEL AND TRANSFORMATIONS

At the heart of the proposed approach is the Quality Model (QM). This is used by Vaucher et al. The Quality Model provides an assessment of the quality of a web page based on a set of metric values extracted from the page. Vaucher et al. define a transformation as a simple modification to a web application that corresponds to a well-defined activity [5]. These transformations impact the quality of the web page and identifying the best sequence of transformations for improving the quality is a problem that cannot be solved in polynomial time [13]. Heuristics can be used for the purpose [7]. For this reason, Vaucher et al use simulated annealing and the present paper proposes the use of Genetic Algorithms.

IV. GENETIC ALGORITHMS

Genetic Algorithms that mimic the natural process of evolution to uncover solutions to complex problems have been successfully applied across multiple domains to solve various types of problems, not just optimization problems. The skeleton of a genetic algorithm is as follows:

Initialize population of solutions at random

Repeat

Evaluate fitness of each solution

Select 2 solutions at random based on fitness

Cross over the solutions to generate a new offspring

Mutate offsprings at random

Change the population by replacing the less fit solutions with the newly generated offsprings

Until termination-criterion-is-met

The challenges to be handled when applying genetic algorithms to solve problems include the selection of a suitable encoding mechanism for solutions, design of a fitness function that can evaluate the fitness of a solution, selection of the type of cross-over to be applied and mutation mechanism to be employed. Also, to be decided are the number of solutions in a generation and the number of generations.

V. MATHEMATICAL FORMULATION OF THE PROBLEM

Let $TS = (t_1, t_2, ..., t_n)$ denote a sequence of transformations selected from a set $T_{candidate}$ of candidate transformations. Let m denote a vector of metric values. The function apply-transform computes new metric values given a transformation sequence TS.

$$m' = apply - transform(TS, m) = tr\_met_n(tr\_met_{n-1}(...tr\_met_1(m))$$

We are interested in identifying the best sequence of transformations TS_{best} that satisfies the following constraints:

$$\forall (TS_i) : q(apply - transform(TS_{best}, m_0)) > q(apply - transform(TS_i, m_0))$$ and

$$Cost(TS) = \sum_{1}^{n} \cos t(t_i) < W$$

VI. PROPOSED METHODOLOGY

The methodology used is very similar to the one employed by Vaucher et al [5]. 50 Web Pages are considered for the study and the initial navigability of the web pages are known. The 50 pages can be grouped into 3 categories: those having good navigability(more than 75%), those with normal navigability (25% to 75%) and those with bad navigability (below 25%). The navigability is expressed as a number and is computed from a set of metrics given below. Table I. depicts metrics considered in this work and Table II. depicts set of transformation and respective cost used.

TABLE I. METRICS CONSIDERED

Metric	Type
Download Speed	Measure
Ratio of links with titles	Measure
Ratio of links with text	Measure
Presence of a search Engine	Binary
Presence of a Menu	Binary
Visited Links Change Color	Binary

Number of Links in Page	Measure
Link to Home	Binary
Support for Back Button	Binary
Presence of Site Map	Binary

Computation of the values of metrics can be achieved using the html unit tool [6]. The following table shows the set of transformations selected and the cost for each transformation. These were the same as used by [5].

TABLE II. SET OF TRANSFORMATIONS AND THE RESPECTIVE COSTS USED

Transformation	**Cost**
Add/Remove Site Map	3
Add/Remove Search Engine	6
Add/Remove URL	2
Add/Remove Link to Home	1
Add/Remove Menu	6
Enable/Disable back Button Support	1
Add/Remove Visited Color Link	2
Correct Link Text	5

The set of selected transformations is minimal as the objective is to show that the proposed methodology is capable of identifying the best sequence of transformations to improve navigability. This set can be extended to include any number of transformations.

The fitness function used for the genetic algorithm is the same as used by [5]:

$$fitness(s) = \begin{cases} 0.5 + Nav\_\mathrm{Im}pr * 0.5 \text{ if } \mathrm{cost(s)} \le \mathrm{W} \\ Nav\_\mathrm{Im}pr * 0.5 \quad \text{ if } \mathrm{cost(s)} > \mathrm{W} \end{cases} \tag{1}$$

Solutions where encoded as chromosomes containing 8 genes with each gene representing the position of a particular transformation in the sequence. Single point cross over was used and the number of generation was capped at 150.

VII. RESULTS

The following table shows the navigability improvement achieved with the 3 categories of web pages with transformation costs of 0%, 25%, 50% and 100% of the total cost. Table III. depicts navigability of pages with costs of 0%, 25%, 50% and 100%. Figure 1. depicts navigability improvement of pages in the 3 categories Bad, Normal and Good.

TABLE III. NAVIGABILITY OF PAGES WITH COSTS OF 0%,25%,50% AND 100%

Category	**0%**	**25%**	**50%**	**75%**	**100%**
Bad	12%	69%	84%	89%	93%
Normal	62%	74%	85%	88%	92%
Good	82%	87%	90%	92%	92%

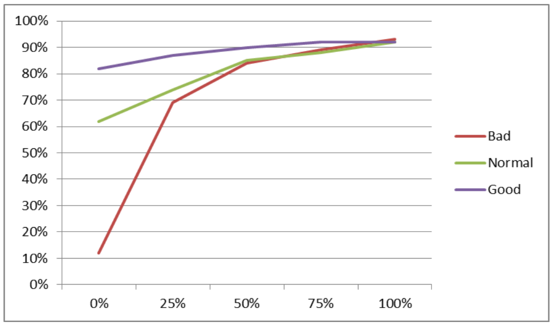

Figure 1. Navigability Improvement of Pages in the 3 categories

As can be seen from the results, the navigability improvement of web pages in the “bad” category tends to be very high as the cost moves from 0% to 25% of the total cost. This is because pages with bad navigability have more number of possible transformations that can improve navigability. The navigability improvement also appears to saturate at more or less the same values of costs for all the 3 categories.

The following table shows the results of comparison of the time needed by the proposed method to find out the best sequence of transformations with those used by Simulated Annealing. Table IV depicts comparison of execution times of the proposed method and simulated annealing and Figure 2. Shows comparison of execution time with SA.

TABLE IV. COMPARISON OF EXECUTION TIMES OF THE PROPOSED METHOD AND SIMULATED ANNEALING

Number of Transformations Considered	**Execution Time (in seconds) with GA**	**Execution Time (in seconds) with SA**
8	6	7
20	8	8
40	13	22
70	18	30
100	27	42

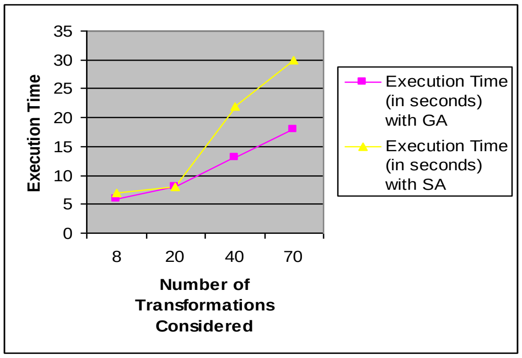

Figure 2. Comparison of Execution Time with SA

As can be observed, the speed improvement brought about by the proposed approach tends to be more visible with increasing number of transformations considered. This is a positive aspect of the proposed method as in real time the number of transformations to be considered will be huge.

VIII. CONCLUSION AND FUTURE WORK

With the increasing number of web sites being developed, the quality of the engineered web sites assumes tremendous importance. In this regard, the paper proposed the use of GA for identifying the best sequence of transformations to improve the navigability of web pages. Improving the quality of web sites is a challenging problem faced by the IT industry at present and the proposed research is a step in the direction if improving quality by using GA. The proposed method is found to yield significant improvements in navigability and performs faster than the existing methods. As a pert of future work, we plan to work on other aspects of quality as well.

REFERENCES

[1] J. Nielsen, H. Loranger," Prioritizing Web Usability", New Riders, Berkeley, Calif, 2006.

[2] C. Boldyruff, P. Warren, C. Gaskell, A. Marshall, "Web-SEM Project: Establishing effective Web Site Evaluation Metrics", Proc. 2nd International Workshop on Web Site Evolution, Washington DC, IEEE Computer Society, 2000.

[3] L. Olsina, G. Rossi, A. Garrido, D. Distante, G. Canfora, "Incremental Quality Improvement in Web Applications Using Web Model Refactoring", Proc. Web Information Systems Engineering Workshops, Springer Berlin / Heidelberg, 2007.

[4] Vaucher, Stephane, Boclinville, Samuel, Sahraoui, Houari, Habra, Naji, "Recommending Improvements to Web Applications using Quality-Driven Heuristic Search", Proc. 10th International Conference on Web Information System Engineering, Poland, 2009.

[5] R.E. Fairley, "Software Engineering Concepts", Tata McGraw Hill, 2001.

[6] R.S. Pressman, "Software Engineering – A Practitioner's Approach", Fifth Edition, 2001.

[7] J.A. Ferland, D.Costa, "Heuristic search methods for combinatorial programming problems", Technical report, Universit_e de Montr_eal, 2001.

[8] M. Haydar, G. Malak, H. Sahraoui, A. Petrenko, S. Boroday, "Anomaly Detectionand Quality Evaluation of Web Applications", Handbook of Research on Web Information Systems Quality, IGI Global , 2008.

[9] E.R. Harold," Refactoring HTML: Improving the Design of Existing Web Applications", 2008.

[10] Karunesh Gupta, Manish Shrivastava, "Web Usage Data Clustering Using Improved Genetic Fuzzy C-Means Algorithm", International Journal of Advanced Computer Research, 2012.

[11] J. Han, M. Kamber, "Data Mining: Concepts and Techniques", Second Edition, Morgan Kaufmann Publishers, 2006.

[12] S B Patil SachinChavan, PreetiPatil, Sunita R Patil," High Quality Design To Enhance and Improve Performance of Large Scale Web Applications",IJCET, 2012.

[13] F. Ricca, P. Tonella, I.D. Baxter, "Restructuring web applications via transformation rules". SCAM, 2001.