Research Paper Computer Science

The Anatomy of Offline Search Engine

S.Gowri Research Scholar Dept of Information Technology, Sathyabama University, Chennai

Dr.G.S.Anandha Mala Dept of Computer Science Engineering, Easwari Engineering College, Chennai

ABSTRACT

There are a plenty of problems when the present searching techniques are been chosen. In this paper we present the Reverse Backward Window Match (RBWM), a prototype for offline search engine which has a faster process time and search time compared to the other search engines. This search engine was idealized from the combination of 2 other existing search engines, which are the Reverse Factor algorithm (RF) and the Backward Non-Deterministic Dawg Matching Algorithm (BNDM). There are enormous researches going on in this field of search process where novel techniques are formulated day-by-day. The proliferation of various types of data would result in difficulty for formulating a common search engine. This paper provides an in-depth description of our offline search engine by addresses the question of how to build a practical offline system which can exploit the additional information present in text.

KEYWORDS Offline Search engines, BNDM, RF.

INTRODUCTION

For information Retrieval the web was a challenge previously, but now the offline search engines are being more challenging, as the amount of information in archive data is growing rapidly. The usage of search in offline is only for searching of folder by comparing of the folder name. In our framework we not just compare the names but the comparison of the content of the folder will also be covered. Low quality matches are retrieved by the traditional search engines which does not even cover all the esoteric topics. There aren't many offline search engines formulated. We have built an offline search engine which addresses at least a few of the problems of existing systems. It makes the use of the binary structure of the query to evaluate and search faster, for provide much higher quality search results.

RELATED WORKS

There are a Plenty of search techniques which are generated by many researchers a few of which are stated below:

Vector distribution, a novel four-step search (4SS) algorithm with center-biased examination point pattern for quick block motion estimation is proposed by Lai-Man Po and Wing-Chung Ma [5]. Halfway-stop technique was used in the new algorithm with searching steps of 2 to 4 and the total number of examination points is varied from 17 to 27. Simulation consequences show that the proposed 4SS executes better than the well-known three-step search and has similar performance to the new three-step search (N3SS) in terms of motion compensation errors. In addition, the 4SS also reduces the worst-case computational necessity from 33 to 27 search points and the average computational necessity from 21 to 19 search points as contrasted with N3SS.

Chun-Ho Cheung and Lai-Man Po [4] proposed two cross-diamond-hexagonal search (CDHS) algorithms, which vary from each other by their sizes of hexagonal search patterns. These algorithms essentially utilize two cross-shaped search patterns successively in the very commencement steps and switch using diamond-shaped patterns. To additional decrease the examination points, two pairs of hexagonal search patterns are proposed in combination with candidates found positioned at diamond corners. Experimental results show that their proposed CDHSs execute faster than the diamond search (DS) by about 144% and the cross-diamond search (CDS) by about 73%, whereas similar prediction quality is still maintained.

Traditional heuristic search algorithms use the ranking of states that a heuristic method offers to guide the search. The objective of Levi H.S. Lelis [3] is to improvise suboptimality and runtime of search algorithms when only fragile heuristics exist—they presented Stratified Tree Search (STS), a suboptimal heuristic search algorithm that utilizes a heuristic to partition the state space to direct the search. They call this partition a type framework. STS assumes that nodes of the similar type will lead to resolutions of the identical cost. Thus, STS increase only one node of each type in every level of search. They had showed in common STS offers a good tradeoff among solution quality and search speed by changing the size of the type system. However, in some cases, STS might not supply a fine alteration of this tradeoff. They presented a variant of STS, Beam STS (BSTS), which permit one to create fine alterations of this tradeoff. BSTS merge the ideas of STS with those of Beam Search. Their empirical results in benchmark domains show that both STS and BSTS can find solutions of minor suboptimality in less time than typical heuristic search algorithms for discovering suboptimal results.

Dennis Weyland [2] stated that in current years a lot of new (mostly naturally inspired) search heuristics have been proposed. Amongst those approaches is Harmony Search. After its introduction in 2000, optimistic results and enhancements over traditional approaches have been reported in numerous publications. They gave a review about the improvements of Harmony Search in the previous decade. Furthermore we execute a precise examination of this approach. In particular we evaluate Harmony Search to the well known and established search heuristic called Evolution Strategies. It turns out that Harmony Search is a special case of Evolution Strategies. They gave an convincing proof for the thesis that research in Harmony, even though certainly conducted with the best of intention, is primarily misguided, marred by a preoccupation with retracing paths previously well traveled, and they conclude that future investigation attempt might better be loyal to more capable areas. The overarching query is how such a technique could become imprecisely depicted as a important innovation without tackling a reputable challenge of its content or credentials. They scrutinize potential answers to this query, and implications for evaluating other evidently pioneering procedures, by disclosing the way in which boundaries of

the technique have been methodically unnoticed.

Weixiong Zhang and Moshe Looks [1] presented and investigated a novel technique for the Traveling Salesman Problem (TSP) that include backbone data into the well-known and extensively useful Lin-Kernighan (LK) local search family of algorithms for the problem. They believed how heuristic backbone data can be acquire and develop techniques to make biased local perturbations in the LK algorithm and its alternatives by developing heuristic backbone data to enhance their effectiveness. They presented wide investigational outcomes, utilizing big examples from the TSP Challenge group and real-world examples in TSPLIB, presenting the important development that the novel function can present over the original algorithms.

SEARCH ENGINE

System features

- This algorithm is the combination of the Reverse Factor algorithm and Backward Non-Deterministic Dawg Matching Algorithm.
- Bit-parallelism simulation is employed of the suffix automaton of x^R and the conversation of string to bits is been chosen from the concept of reverse factoring algorithm.
- If the pattern length should not extend the machine's memory-word size, the algorithm is exceptionally effective if these criteria are satisfied. As a result dynamic variation in the size of the window is done which is been altered with the help of the word size.

Description of RBWM

- The search is executed in a window which is used to move forward incrementally without looping.
- When there is no active state (that is the flag value altered after permutation trails) the shift of window occurs, which is same as in the concept of BDN, since no looping process is done.
- The Search status updating formula is:

$$D_j = \left(D_{j-1} AND\ B[S_j]\right) \ll 1$$

- Explanation of the update formula
- No input symbol can re-activate any machine, instead only an active machine can receive a new symbol.
- The next input symbol regulates on which active machine it will be feasible to try process it
- The bit mask estimates which machine can execute the input symbol (those having the corresponding 1 in the bit mask, i.e. b_i=1)
- To remain active, a machine must previously be active (d_i=1) and the bit mask for that machine must be 1 (therefore AND in the formula)
- The backward scan of the window reflects the left shit by one.
- Compared to Shift-And and Shift-Or the bit mask has reverse order, since the search is backwards.

Example: if *w*=bbaac, B[a]=00110, B[b]=11000, B[c]=00001

Algorithm

- A window with a dynamic memory allocation is set for the purpose of search (the dynamic allocation is done cause the size of the window depends on the word for comparison)
- The word search is made without looping process; instead the window is moved when the character wise comparison does not match the hit word.
- When the word is found in the clusters, the count value to the searched word is also set on increment and the resultant is displayed along with the word count in a sorted manner.

Table B of size ASIZE in this clustering algorithm is initialized for separate character c, for which a bit mask is saved. If and only if x_i=c the mask in B_c is set. In a word $d=d_{m-1} .. d_0$ the search state is retained, where the machine word size is greater than or equal to the pattern length m.

If an only if x[m-i .. m-1-i+k]=y[j+m-k .. j+m-1] the bit d_i at iteration k is set. Variable d is set to 1^{m-1} at iteration 0. To update d follows d'= (d & B[y_j]) << 1 formula. If and only if, after iteration m, it holds d_{m-1}=1, then there is a match of the key word which is on hit.

The algorithm has imitated a prefix of the pattern in the current window position j, whenever d_{m-1}=1. The shift to the next position is given when the longest prefix is imitated.

CONCLUSION

In this paper the searching process is the newly introduced and implemented, which is formulated utilizing the Backward Non-Deterministic Dawg Matching Algorithm and the Reverse Factor algorithm. These algorithms were chosen in particular due to their features which are highlighting, such as the Window concept of the BNDM Algorithm and the reverse comparison of the RF algorithm. The process time would get decreased with the help of the Window concept and the relevancy in search results would increase due to reverse comparison concept. Additional the concept of the binary conversion of the query would assist in the process and search time to get reduced drastically.

REFERENCES

[1] Weixiong Zhang and Moshe Looks, "A novel local search algorithm for the traveling salesman problem that exploits backbones", Proceedings of the 19th international joint conference on Artificial intelligence IJCAI '05, pp. 343-348, 2015. | [2] Dennis Weyland, "A Rigorous Analysis of the Harmony Search Algorithm: How the Research Community can be Misled by a "Novel" Methodology", International Journal of Applied Metaheuristic Computing archive, Volume 1 Issue 2, pp. 50-60, April 2010. | [3] Levi H.S. Lelis, Sandra Zilles and Robert C. Holte, "Stratified Tree Search: A Novel Suboptimal Heuristic Search Algorithm", Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems AAMAS '13, pp. 555-562, 2014. | [4] Chun-Ho Cheung and Lai-Man Po, "Novel Cross-Diamond-Hexagonal Search Algorithms for Fast Block Motion Estimation", IEEE Transactions on Multimedia, Volume:7 , Issue: 1, pp- 16 – 22, Feb 2005. | [5] Lai-Man Po and Wing-Chung Ma, "A Novel Four-Step Search Algorithm for Fast Block Motion Estimation", IEEE Transactions on Circuits and Systems for Video Technology, Volume:6 , Issue: 3, pp. 313 – 317, Jun 1996. |