

Profile of Bucket Tries

Mehri Javanian*

Faculty of Sciences, University of Zanjan, Zanjan, Iran

Received: 5 Aug. 2018 Accepted: 28 October 2018

Extended Abstract

Paper pages (169-182)

Introduction

Tries (from *retrieval*) are one of the most practical data structures with a tree construction in computer science. Tries store string data in leaves of tree. They are often used to store such data so that future retrieval can be made efficient. For example, tries are widely used in algorithms for automatically correcting words in texts. The number of nodes of the same type, which are at the same distance from the root of a rooted tree, is called profile. The analysis of the profile of a tree is of great importance. Because many of the parameters of a rooted tree can be expressed in terms of its profile. Although profiles represent one of the most fundamental parameters of tries, they have hardly been studied in the past. A generalized version of trie is called bucket trie where each leaf or bucket has a storage capacity of more than one string. Random trie is obtained by defining a random grow rule for trie. We present a detailed study of the limit behavior of the profiles in a random bucket tries.

Material and methods

In this paper, the methods we apply to derive recurrences satisfied by the expected profiles and to solve them asymptotically for all possible ranges of the distance from the root, are based on the use of Poissonization, Mellin transform, recurrence equations, generating functions, singularity analysis and saddle-point method.

Results and discussion

Here, as the number of stored strings in a random bucket trie increases, we investigate the asymptotic expectation, variance and limiting distribution of each of the two internal and bucket profiles (i.e. the number of bucket nodes or leaves, and the number of internal nodes or non-leaves which are at the same distance from the root) in random bucket tries. Both the expectation and variances of the two profiles contain periodic functions, and we show those periodic functions are not zero that this point has not been proven in the paper on the ordinary trie. Also, we examine the amount of asymptotic ratio of the expectations of the bucket and internal profiles.

Conclusion

In this research, we will generalize the most important part of the results for the ordinary tries and also provide a proof for an unproven point in those results. More precisely, the purpose of this article is to study the internal profile (the number of non-leaf nodes at distance k from the root) and bucket profile (the number of buckets at distance k from the root) in an important data structure called random bucket trie (random trie based on n data and maximum capacity of

$b \geq 1$ data per each leaf). By some methods in complex analysis, we have shown that for every $\varepsilon > 0$ and some fixed constants α_1 and α_2 , if k is in the range

$$(\alpha_1 + \varepsilon) \log n \leq k \leq (\alpha_2 - \varepsilon) \log n,$$

then the expectations and variances of both profiles contain non-zero periodic functions (the non-zero property of these periodic functions has not been proved for the ordinary tries up to now). We also provide a graph of those periodic functions for $b = 1, 2, 3, 4$ by MAPLE. Then we study the ratio of the two profiles for $b \geq 1$. Finally, by finding the asymptotic expansions of Poisson generating functions for the probability generating functions of profiles and then using the Cauchy integral formula, we obtain the asymptotic expansions for the probability generating functions, which indicate that the limiting distributions of profiles are normal.

Keywords: Bucket tries, Profile; Poissonization, Mellin transform, recurrence equations, generating functions, singularity analysis, saddle-point method.

*Corresponding author: javanian@znu.ac.ir