

AN EFFICIENT MAPPING FOR COMPUTING THE SCORE OF STRING MATCHING ¹

TETSUYA NAKATOH

Computing and Communications Center, Kyushu University
6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan
e-mail: nakatoh@cc.kyushu-u.ac.jp

KENSUKE BABA, DAISUKE IKEDA

Department of Informatics, Kyushu University
744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan
e-mail: {baba,daisuke}@i.kyushu-u.ac.jp

YASUHIRO YAMADA

Kyushu University User Science Institute
6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan
e-mail: yamada@usi.kyushu-u.ac.jp

and

SACHIO HIROKAWA

Computing and Communications Center, Kyushu University
6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan
e-mail: Hirokawa@cc.kyushu-u.ac.jp

ABSTRACT

The authors propose an efficient algorithm to solve the problem of *string matching with mismatches*. For a text of length n and a pattern of length m over an alphabet Σ , the problem is known to be solved in $O(|\Sigma|n \log m)$ time by computing a score by the fast Fourier transformation (FFT). Atallah et al. introduced a randomized algorithm in which the time complexity can be decreased by the trade-off with the accuracy of the estimates for the score. The algorithm in the present paper yields an estimate with smaller variance compared to the algorithm by Atallah et al. Moreover it computes the exact score in $O(|\Sigma|n \log m)$ time.

Keywords: String matching, mismatch, FFT, convolution, randomized algorithm

¹Full version of a submission presented at the *Prague Stringology Conference* (Czech Technical University in Prague, Czech Republic, September 22–24, 2003).