

Hon, Wing-Kai; Lam, Tak-Wah; Shah, Rahul; Thankachan, Sharma V.; Ting, Hing-Fung; Yang, Yilin

Dictionary matching with uneven gaps. (English) [Zbl 1383.68105](#)

Cicalese, Ferdinando (ed.) et al., Combinatorial pattern matching. 26th annual symposium, CPM 2015, Ischia Island, Italy, June 29 – July 1, 2015. Proceedings. Cham: Springer (ISBN 978-3-319-19928-3/pbk; 978-3-319-19929-0/ebook). Lecture Notes in Computer Science 9133, 247-260 (2015).

Summary: A gap-pattern is a sequence of sub-patterns separated by bounded sequences of don't care characters (called gaps). A one-gap-pattern is a pattern of the form $P[\alpha, \beta]Q$, where P and Q are strings drawn from alphabet Σ and $[\alpha, \beta]$ are lower and upper bounds on the gap size g . The gap size g is the number of don't care characters between P and Q . The dictionary matching problem with one-gap is to index a collection of one-gap-patterns, so as to identify all sub-strings of a query text T that match with any one-gap-pattern in the collection. Let \mathcal{D} be such a collection of d patterns, where $\mathcal{D} = \{P_i[\alpha_i, \beta_i]Q_i \mid 1 \leq i \leq d\}$. Let $n = \sum_{i=1}^d |P_i| + |Q_i|$. Let γ and λ be two parameters defined on \mathcal{D} as follows: $\gamma = |\{j \mid j \in [\alpha_i, \beta_i], 1 \leq i \leq d\}|$ and $\lambda = |\{\alpha_i, \beta_i \mid 1 \leq i \leq d\}|$. Specifically γ is the total number gap lengths possible over all patterns in \mathcal{D} and λ is the number of distinct gap boundaries across all the patterns. We present a linear space solution (i.e., $O(n)$ words) for answering a dictionary matching query on \mathcal{D} in time $O(|T|\gamma \log \lambda \log d + \text{occ})$, where occ is the output size. The query time can be improved to $O(|T|\gamma + \text{occ})$ using $O(n + d^{1+\epsilon})$ space, where $\epsilon > 0$ is an arbitrarily small constant. Additionally, we show a compact/succinct space index offering a space-time trade-off. In the special case where parameters α_i and β_i 's for all the patterns are same, our results improve upon the work by A. Amir et al. [Lect. Notes Comput. Sci. 8486, 11–20 (2014; [Zbl 1390.68781](#))]. We also explore several related cases where gaps can occur at arbitrary locations and where gap can be induced in the text rather than pattern.

For the entire collection see [\[Zbl 1314.68012\]](#).

MSC:

[68W32](#) Algorithms on strings

[68P05](#) Data structures

Cited in **2** Documents

Keywords:

[dictionary matching](#); [point enclosure queries](#)

Full Text: [DOI](#)