In previous publications the authors only defined the relative substring error not the absolute substring error. Below both the relative as well as the absolute substring error are defined.

0.1. The substring error: σ

This measure provides additional useability and flexibility to the user. The FireμSat$_2$ software provides the user with the option to specify whether calculations should be based on either the relative or the the absolute substring error. These measures are computed at appropriate points by FireμSat$_2$ and then compared against a user-specified threshold value of the *maximum substring error allowed*, τ. During processing $\sigma \leq \tau$ should always hold.

In line with the guidelines suggested by Benson, [1], the value of σ depends, *inter alia*, on penalties (or weights) allocated by the user to mismatches (p_m), deletions (p_d) and insertions (p_i). In each case, the allowable range is a real number in the interval $[0, 1]$. For a given motif, ρ, and a given substring that has been partitioned into the form $u = \rho u_2 \cdots u_p$, σ on u can be computed either as the absolute or as the relative substring error as follows:

- *The absolute substring error.*

$$\sigma = (n\_d * p\_d) + (n\_i * p\_i) + (n\_m * p\_m)$$

- *The relative substring error.*

$$\sigma = (n\_d * p\_d) + (n\_i * p\_i) + (n\_m * p\_m) - n\_ptre$$

In each case, n_d is the number of deletions in u; n_i is the number of insertions in u and n_m is the number of mismatches in u.

At the time of writing, two implementations of FireμSat$_2$ were made available. The one implementation calculates the absolute substring error while the relative substring error is computed by the other implementation. The user may rely on system default values for the penalties. These are $p\_i = 1.0$, $p\_d = 1.0$ and $p\_m = 0.5$ respectively. Alternatively, specific values may be selected to attain various specific objectives. For example, a penalty weight of 0 may be chosen for one or more of the mutation types, in which case no penalty is assigned to ATREs that derive from that mutation type.

σ is calculated after each detection of a TRE. It is then compared against a user specified threshold value, τ. If $\sigma > \tau$ then execution is terminated and the TR to date is reported on provided that the other conditions, set by the user, hold.

In the case of the relative substring error the value of σ therefore reflects the extent to which the number of ATREs exceeds the number of PTREs, weighted in terms of penalty values associated with mismatches, deletions and insertions.

A restrictive default value has not been allocated to τ. If the user does not enter a value for τ then the substring error, σ, may be arbitrarily large. Note, the relative substring error is similar to the substring error defined for FireμSat$_1$, in [2] and [3]

Whether absolute or relative substring error is computed, the user may assign values to penalties progressively to reveal more and more information as deemed appropriate. For example, TRs that consist only of PTREs and ATREs with mismatches may be requested in a first run. If sufficient TRs are not detected then the deletion penalty values and/or insertion penalty values can be adjusted such

that TRs containing mismatches, deletions and/or insertions will be detected in a second run.

References

1. G. Benson, *Tandem repeats finder*, Nucleic acids research **27** (1999), no. 2, 573 – 580.
2. C. De Ridder, D.G. Kourie, and B.W. Watson, *Fireµsat: an algorithm to detect microsatellites in dna*, Proceedings of the Prague Stringology Conference 2006 (2006).
3. _____, *Meeting the challenge of detecting microsatellites in dna*, South African Institute of Computer Scietists and Information Technologists:Service-oriented Software and Systems (2006).

DNA-algo group, FASTAR-research group, Department of Computer Science, School of Information Technology, University of Pretoria, South-Africa.
  E-mail address: driddc@unisa.ac.za