Prof. Dr. Knut Reinert
Jacob Schulze
René Rahn

Institut für Informatik
AG Algorithmische Bioinformatik

# Algorithmen und Datenstrukturen in der Bioinformatik
# Dritte Programmieraufgabe WS 14

**Abgabe Montag, 1.12., 15:00 Uhr per SVN**

In dieser Programmieraufgabe werden Sie die exakte Stringsuche mithilfe eines Suffixarrays implementieren.

**Aufbau** Schreiben Sie eine Funktion `void construct(vector<unsigned> & sa, string const & text)`, die ein leeres Suffixarray sowie den Text erhält und das Suffixarray konstruiert. Nutzen Sie die naive Konstruktionsmethode mit `std::sort`. Beachten Sie, dass ein Suffixarray niemals Strings speichert, sondern lediglich deren Startpositionen im Originaltext.

- Anders als in den theoretischen Aufgaben benötigen sie kein extra $ am Ende des Textes.
- Um `std::sort` zu nutzen müssen Sie einen sog. Funktor anlegen, der den $<$-Operator für 2 Textpositionen definiert. Beispiele hierfür gibt es im Netz, u.A. unter `http://www.cplusplus.com/reference/algorithm/sort/`.

**Suche** Programmieren Sie die Binärsuche mit der *mlr*-Heuristik in der Funktion `void find(std::vector<unsigned> & hits, string const & query, std::vector<unsigned> const & sa, std::string const & text)`. In dem vector `hits` werden die Anfangspositionen der Treffer im Text gespeichert. Achten Sie auf den `&`-operator, der eine Adresse und nicht die Kopie des Vektors übergibt.

**Aufruf** Das Programm kann in 2 Modi ausgeführt werden. Gibt man nur ein Argument - den Text - an, so soll das Suffixarray zeilenweise ausgegeben werden. Gibt man nach dem Text noch ein oder mehrere Suchwörter als weitere Kommandozeilenparamter an, dann sollen zeilenweise die Suchwörter, sowie, durch Leerzeichen getrennt, die Liste der Positionen der Treffer im Text ausgegeben werden. Bei nicht korrektem Aufruf soll das Programm `unexpected input` ausgeben und den return code -1 zurück geben.

Beispiele für beide Modi:

```
./aufgabe3 "banana"
5
3
1
0
4
2
```

```
./aufgabe3 "exact search using suffix arrays" "s" "horspool" "g suf"
s: 6 14 19 31
horspool:
g suf: 17
```

**Praktikumshinweise** **WICHTIG!** Achten Sie auf die korrekte Ausgabe. Die Ausgabe im 2. Modus entspricht dem Format: *suchwort: treffer_1 treffer_2 ...*. Sollte kein Treffer gefunden werden, wird nur das Suchwort und der Doppelpunkt ausgegeben. Beachten Sie die Hinweise unter `https://www.mi.fu-berlin.de/w/ABI/AlDaBiWS14Praktikum`.