Implementing bioinformatics algorithms in \TeX

**Gotoh** package, a case study

Takuto ASAKURA (wtsnjp)

The University of Tokyo

TUG@Bacho\TeX 2017
Biological sequences

DNA, RNA, Amino acids, etc.

Biologists want to know the degree of similarity among 2 or more sequences.
Pairwise sequence alignment

The problem

**Input:** Two biological sequences

\[ A \equiv a_1 a_2 a_3 \ldots a_m, \quad B \equiv b_1 b_2 b_3 \ldots b_n \]

where \( a_i \) and \( b_j \) are chosen from a finite alphabet, e.g. \{A, T, G, C\}.

**Output:** An alignment between \( A \) and \( B \).

**Examples**

```
     bachotex  context  bioinformatics
---------- ------ ------------
|**|*|***|**|**|  |**|**|**|
```

```
     bachotek  ---tex-  bi-----blat-ex
---------- ------ ------------
    |*| **|**|  |**|**|**|
```
Longest Common Subsequence (LCS)

LCS problem

▶ Want to get the LCS of $A$ and $B$
▶ A simplest form of sequence alignment
▶ Score 1 for matches and 0 for gaps

The solution

$$s_{i,j} = \max \begin{cases} s_{i-1,j} \\ s_{i,j-1} \\ s_{i-1,j-1} + 1 \end{cases}$$
The Gotoh algorithm: DP

Sequence alignment has a slightly more complex scoring scheme.

Example

\[
G^\text{NcB} = 1, \quad G^\text{CMG}^\text{NcB} = -1, \quad g(l) = -d - (l - 1)e
\]

The algorithm

Sequence alignment in \(O(mn)\) time:

\[
M_{i+1,j+1} = \max \left\{ M_{ij}, I_{xij}, I_{yij} \right\} + c_{ai}b_j
\]

where

\[
I_{x_i+1,j} = \max \left\{ M_{ij} - d, I_{xij} - e, I_{yij} - d \right\},
\]

\[
I_{y_i,j+1} = \max \left\{ M_{ij} - d, I_{yij} - e \right\}.
\]
The Gotoh algorithm: trace back

Start at maximum entry, trace back to first entry.

\[
\begin{array}{cccccc}
G & A & G & A \\
0 & -\infty & -\infty & -\infty & -\infty & -\infty \\
G & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & -\infty & 1 & -8 & -7 & -10 \\
A & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & -\infty & -8 & 2 & -7 & -6 \\
C & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & -\infty & -9 & -7 & 1 & -6 \\
T & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & -\infty & -10 & -8 & -6 & 0 \\
A & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & -\infty & -11 & -7 & -7 & -5
\end{array}
\]

GACTA
GA-GA
What I did is . . .

\LaTeX\ package

\texttt{\LaTeX} A Turing Machine
\LaTeX{} Widely used for typesetting papers

+ 

The Gotoh algorithm

- Can be written in short code
- Calculated with limited range of numbers
- Produces visual results
The **Gotoh** package

**Usage**

- \Gotoh\{\textit{sequence A}\}\{\textit{sequence B}\}
  - Executes the algorithm
  - Returns the results to specified CSs
- \GotohConfig\{\textit{key-value list}\}
  - Setting various parameters
  - e.g. algorithm parameters, CSs to store results

**Example**

**Input:**

\Gotoh{ATCGGCGCACGCGGGA}\{TTCCGCCCACA\}
\texttt{\texttt{GotohResultA}}
\texttt{\texttt{GotohResultB}}

**Output:**

ATCGGCGCACGCGGGA
TTCCGCCCACA......A
Combining with TeXshade

The TeXshade package

- A part of BioTeX, produced by Eric Beitz
- Shading and labeling preprocessed alignments
- Can be used to format the outputs of Gotoh

Example

\newcommand{\PrintAlignment}[3][\relax]{% 
  \Gotoh{#2}{#3}% 
  \immediate\openout\FASTAfile=\jobname.fasta 
  \writeFASTA{\> Seq 1^^J\GotohResultA}% 
  \writeFASTA{\> Seq 2^^J\GotohResultB}% 
  \immediate\closeout\FASTAfile 
  \texshade{\jobname.fasta}\#1\endtexshade}

Let me show you a demonstration!
Features and future

Advantages
The \texttt{Gotoh} package is:

\begin{itemize}
  \item simple to use
  \item long-lasting
  \item cross-platform
\end{itemize}

Future work

\begin{itemize}
  \item Preparing the documentation
  \item Uploading to CTAN
  \item Adding some functions such as:
    \begin{itemize}
      \item showing edit graphs
      \item calculating multiple alignment ($\geq 3$ sequences)
    \end{itemize}
\end{itemize}
Conclusion

Algorithms in any field which are:
  ▶ often used for creating documents
  ▶ easy to implement
are worth implementing in \TeX{}.

Example
diff function for \texttt{listings}

Thank you & Happy \TeX{}ing!!