FACULTY OF SCIENCE,
TECHNOLOGY
AND MEDICINE

Longest Common Subsequences

Seminar 2

Joris LIMONIER

Supervised by George KERCHEV

May 31, 2021

Table of Contents

1. Introduction

1.1 What are LCS ?

1.2 Why are we interested in LCS ?

2. How to find LCS ?

2.1 Step 1: Building the table

2.2 Step 2: Crawling back up the table

3. Data analysis of LCS results

3.1 Subsection 1

1. Introduction

1. Introduction

1.1 What are LCS ?

What are LCS ?

Notation

"LCS" = Longest Common Subsequence(s)

What are LCS ?

Notation

"LCS" = Longest Common Subsequence(s)

Example 1

S_1 : A B A B B

What are LCS ?

Notation

"LCS" = Longest Common Subsequence(s)

Example 1

$$\begin{array}{lccccc} S_1: & A & B & A & B & B \\ S_2: & A & A & B & A & B \end{array}$$

What are LCS ?

Notation

"LCS" = Longest Common Subsequence(s)

Example 1

S_1 : **A** B A B B

S_2 : **A** A B A B

What are LCS ?

Notation

"LCS" = Longest Common Subsequence(s)

Example 1

S_1 : **A** **B** A B B

S_2 : **A** A **B** A B

What are LCS ?

Notation

"LCS" = Longest Common Subsequence(s)

Example 1

S_1 : **A** **B** **A** B B

S_2 : **A** A **B** **A** B

What are LCS ?

Notation

"LCS" = Longest Common Subsequence(s)

Example 1

S_1 : **A** **B** **A** **B** B

S_2 : **A** A **B** **A** **B**

What are LCS ?

Notation

"LCS" = Longest Common Subsequence(s)

Example 1

S_1 : **A** **B** **A** **B** B

S_2 : **A** A **B** **A** **B**

$\Longrightarrow$ The LCS between S_1 and S_2 is **A B A B**

What are LCS ?

Notation

"LCS" = Longest Common Subsequence(s)

Example 1

$$S_1 : \quad \mathbf{A} \quad \mathbf{B} \quad \mathbf{A} \quad \mathbf{B} \quad \mathrm{B}$$
$$S_2 : \quad \mathbf{A} \quad \mathrm{A} \quad \mathbf{B} \quad \mathbf{A} \quad \mathbf{B}$$

$\Longrightarrow$ The LCS between S_1 and S_2 is **A B A B**

NB: LCS may not be unique, A A B B also works.

Example 2

What is the LCS of the following sequences ?

Example 2

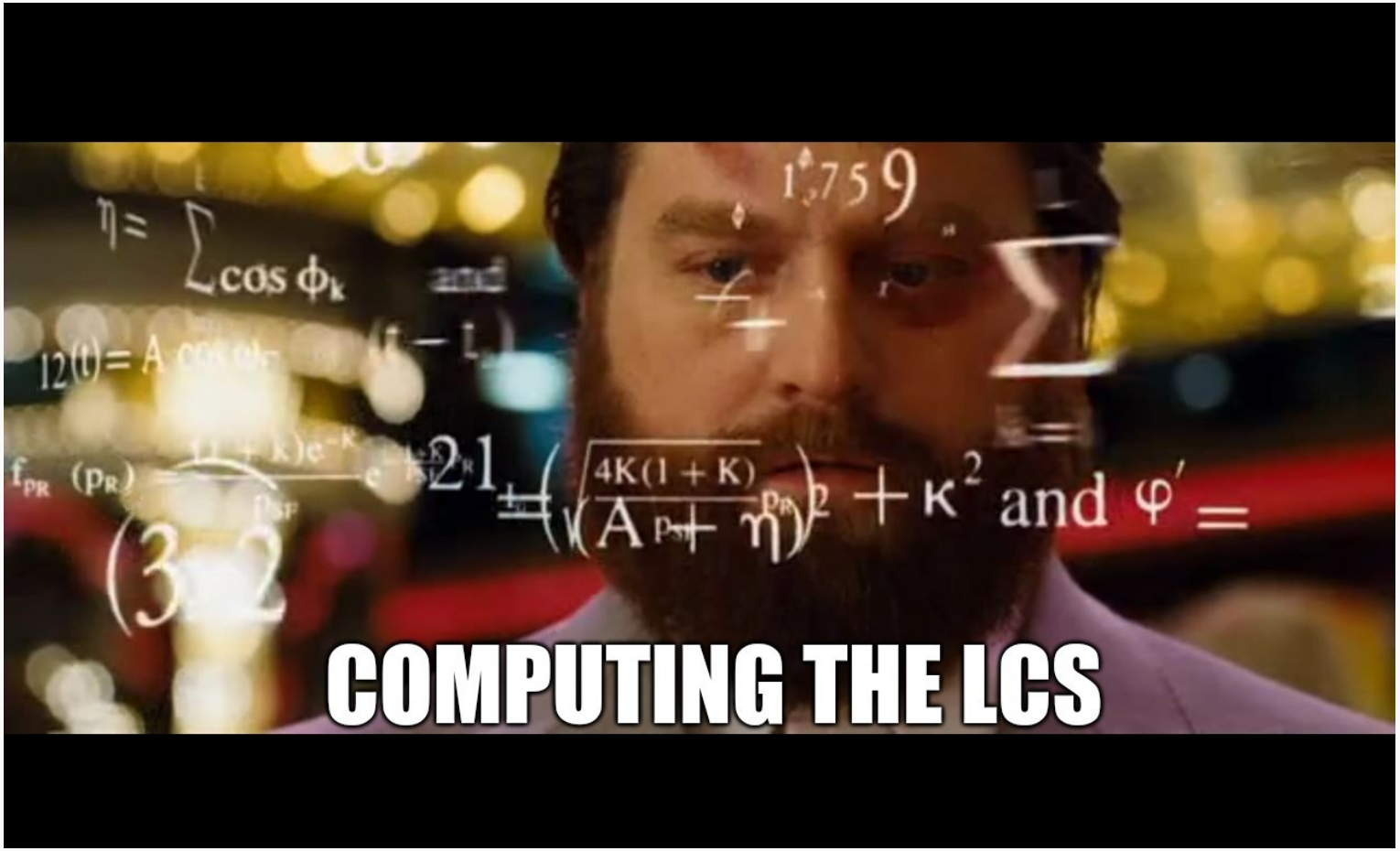
What is the LCS of the following sequences ?

S_3 : AABBAABAAABABAAABAAABBABAAABAAABAAA
AAAABBABBBBAAABABBAABBAABBBBBAAAABA
BBABAAAABABAABBBABABBBBBAAABBBAABBB
AABAABBABABAABABBBBBBAABBBBBBBAAAAAB
AABAAAAABAABAABAAABBABBBBABBAAAABBB

S_4 : BABBBABAABAABBBABBABBBBBBBABABAAABB
BBABBABBABBBABBBBABBABBABABABBAABABA
BAABABAAAABABBABABBAAAABABBAABABABB
BABBBBBBAABAAABBABBBBAAAAABBBBBAAAB
ABBAAAABBBABABAABBABBBAABABBBABAABA

Example 2

What is the LCS of the following sequences ?



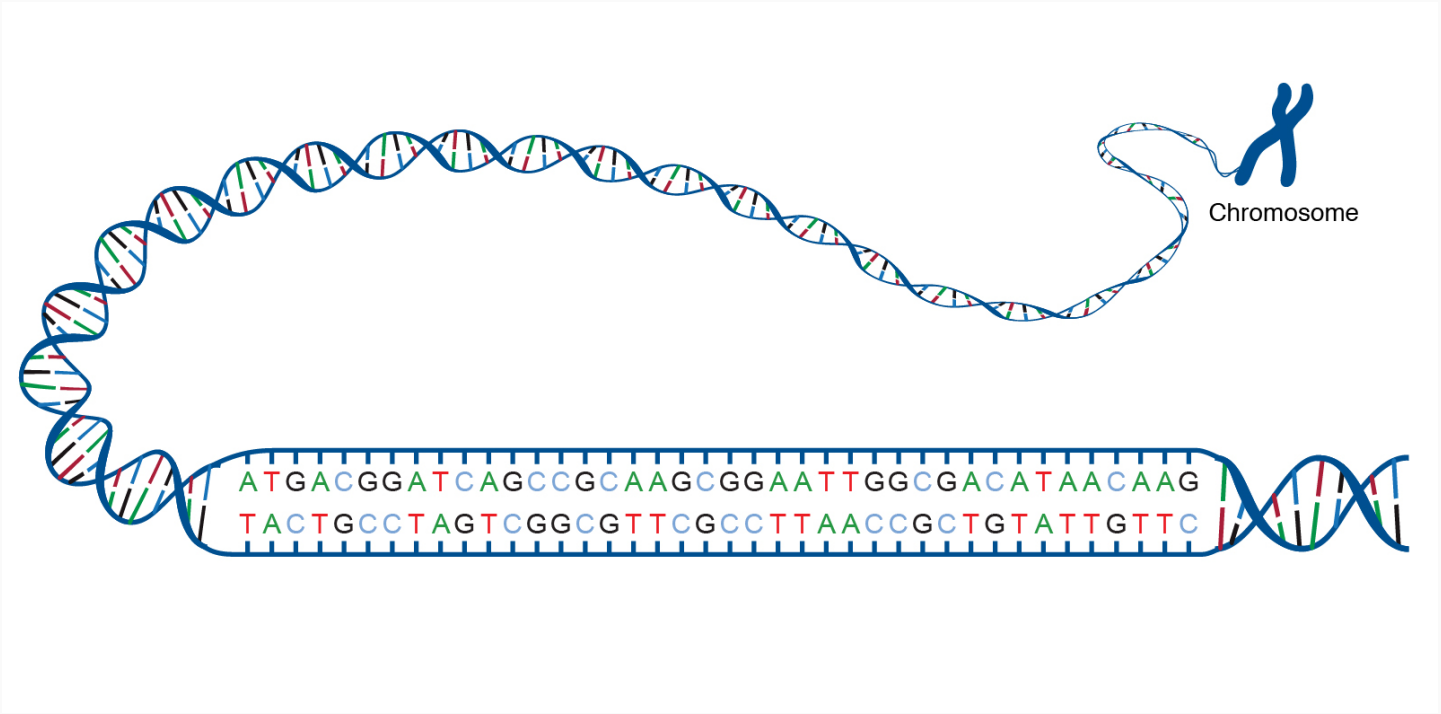
1. Introduction

1.2 Why are we interested in LCS ?

Applications

Applications:

- Bioinformatics: Compare sequences of nucleotides (DNA)

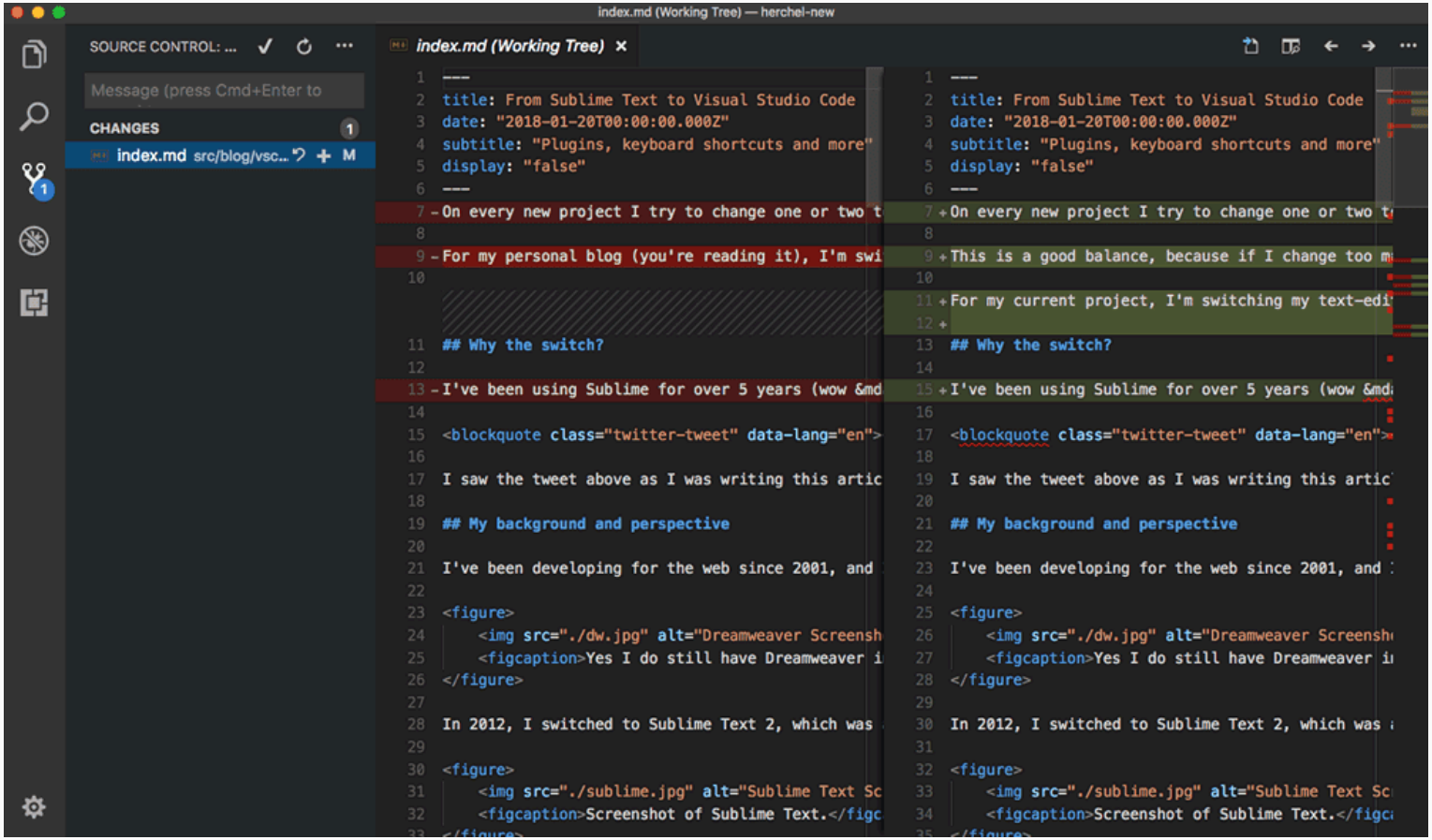


Applications:

- Bioinformatics: Compare sequences of nucleotides (DNA)
- Natural Language Processing: Compare texts

Applications:

- Bioinformatics: Compare sequences of nucleotides (DNA)
- Natural Language Processing: Compare texts
- Computer Science: Spot differences in texts



2. How to find LCS ?

2. How to find LCS ?

2.1 Step 1: Building the table

2. How to find LCS ?

2.2 Step 2: Crawling back up the table

3. Data analysis of LCS results

3. Data analysis of LCS results

3.1 Subsection 1

Thank you