Dynamic Programming

Motivation

- Suppose we put a pair of rabbits in a place surrounded on all sides by a wall.
- How many pairs of rabbits can be produced from that pair in a year if it is supposed that every month each pair begets a new pair which from the second month on becomes productive?
- Resulting sequence is
  1, 1, 2, 3, 5, 8, 13, 21, 34, 55, . . .
- Each number is the sum of the two preceding numbers.

Leonardo Fibonacci Pisano

- This problem was posed by Leonardo Pisano, better known by his nickname Fibonacci (son of Bonacci, born 1170, died 1250).
- This problem and many others were in posed in his book *Liber abac*i, published in 1202.
- The book was based on the arithmetic and algebra that Fibonacci had accumulated during his travels.

Leonardo Fibonacci Pisano

- The book, which went on to be widely copied and imitated, introduced the Hindu-Arabic place-valued decimal system and the use of Arabic numerals into Europe.
- The rabbits problem in the third section of Liber abaci led to the introduction of the Fibonacci numbers and the Fibonacci sequence for which Fibonacci is best remembered today.

Leonardo Fibonacci Pisano

- This sequence, in which each number is the sum of the two preceding numbers, has proved extremely fruitful and appears in many different areas of mathematics and science.
- The *Fibonacci Quarterly* is a modern journal devoted to studying mathematics related to this sequence.

Fibonacci Sequence

- The Fibonacci numbers F_n are defined as follows:

$F_0 = 0$
$F_1 = 1$
$F_n = F_{n-1} + F_{n-2}$

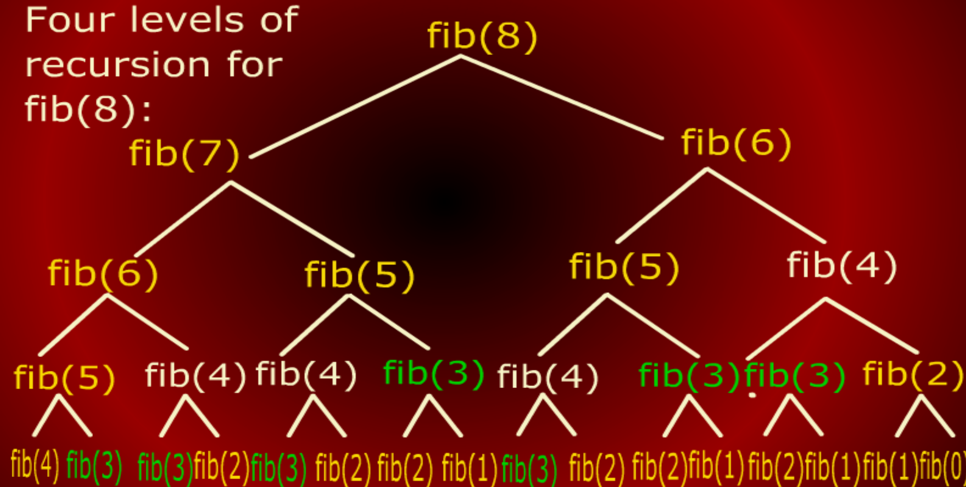
Fibonacci Number

The recursive definition of Fibonacci numbers gives us a recursive algorithm for computing them:

```
FIB(n)
  if (n < 2)
    then return n
    else return FIB(n - 1)+FIB(n - 2)
```

Fibonacci Number: Recursive Calls

Four levels of recursion for fib(8):



Fibonacci Number: Recursive Calls

- A single recursive call to fib(n) results in one recursive call to fib(n - 1), two recursive calls to fib(n - 2), three recursive calls to fib(n - 3), five recursive calls to fib(n - 4) and, in general, F_{k-1} recursive calls to fib(n - k)
- For each call, we're recomputing the same fibonacci number from scratch.

Fibonacci Number: Recursive Calls

- We can avoid this unnecessary repetitions by writing down the results of recursive calls and looking them up again if we need them later.
- This process is called *memoization.*

Fibonacci Number: Memoization

```
MEMOFIB(n)
if (n < 2)
    then return n
if (F[n] is undefined)
    then F[n] ← MEMOFIB(n – 1)
                + MEMOFIB(n – 2)
return F[n]
```

Fibonacci Number: Recursive Calls

- If we trace through the recursive calls to MEMOFIB, we find that array F[] gets filled from bottom up.

```
MEMOFIB(n)
if (n < 2)
  then return n
if(F[n] is undefined)
  then F[n] ←
      MEMOFIB(n - 1)
     +MEMOFIB(n - 2)
return F[n]
```

Fibonacci Number: Recursive Calls

to MEMOFIB, we find that array F[] gets filled from bottom up.

- i.e.,first F[2], then F[3], and so on, up to F[n].

```
MEMOFIB(n)
if (n < 2)
  then return n
if(F[n] is undefined)
  then F[n] ←
      MEMOFIB(n - 1)
     +MEMOFIB(n - 2)
 return F[n]
```

Fibonacci Number: Recursive Calls

- We can replace recursion with a simple for-loop that just fills up the array F[] in that order.

```
MEMOFIB(n)
if (n < 2)
  then return n
if(F[n] is undefined)
  then F[n] ←
      MEMOFIB(n - 1)
     +MEMOFIB(n - 2)
 return F[n]
```

Fibonacci Number: Iterative Algorithm

This gives us our first explicit dynamic programming algorithm.

```
ITERFIB(n)
    F[0] ← 0
    F[1] ← 1
    for i ← 2 to n
    do
        F[i] ← F[i – 1] + F[i– 2]
    return  F[n]
```

Fibonacci Number: Iterative Algorithm

- This algorithm clearly takes only O(n) time to compute F_n.
- By contrast, the original recursive algorithm takes

$$\Theta(\Phi^n),\ \Phi = \frac{1+\sqrt{5}}{2} \approx 1.618.$$

- ITERFIB achieves an exponential speedup over the original recursive algorithm.

Dynamic Programming

- Dynamic programming is essentially recursion without repetition.
- Developing a dynamic programming algorithm generally involves two separate steps:

Dynamic Programming

- **Formulate problem recursively.** Write down a formula for the whole problem as a simple combination of answers to smaller subproblems.
- **Build solution to recurrence from bottom up**. Write an algorithm that starts with base cases and works its way up to the final solution.

Dynamic Programming

- Dynamic programming algorithms need to store the results of intermediate subproblems.
- This is often *but not always* done with some kind of table.

Edit Distance

- Introduced by Levenshtein in 1966.
- Definition: Minimum number of edit operations to transform one string to another
- Possible edit operations
  - Symbol insertion (I)
  - Symbol deletion (D)
  - Symbol substitution (S)

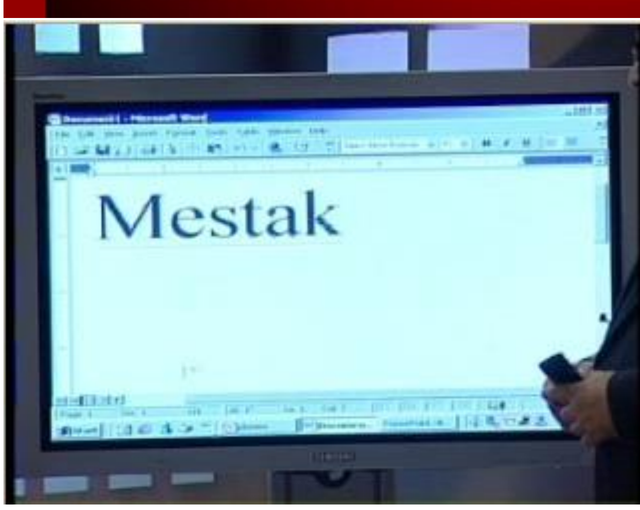
Edit Distance

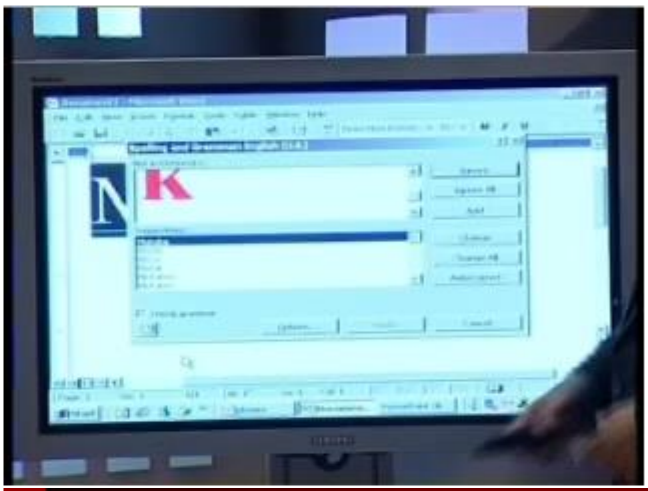
- For example, the edit distance between FOOD and MONEY is at most four:

<u>F</u>OOD → MO<u>O</u>D → MON‸D
→ MONE<u>D</u> →MONEY

Edit Distance: Applications

Spelling Correction
If a text contains a word that is not in the dictionary, a ‘close’ word, i.e. one with a small edit distance, may be suggested as a correction.

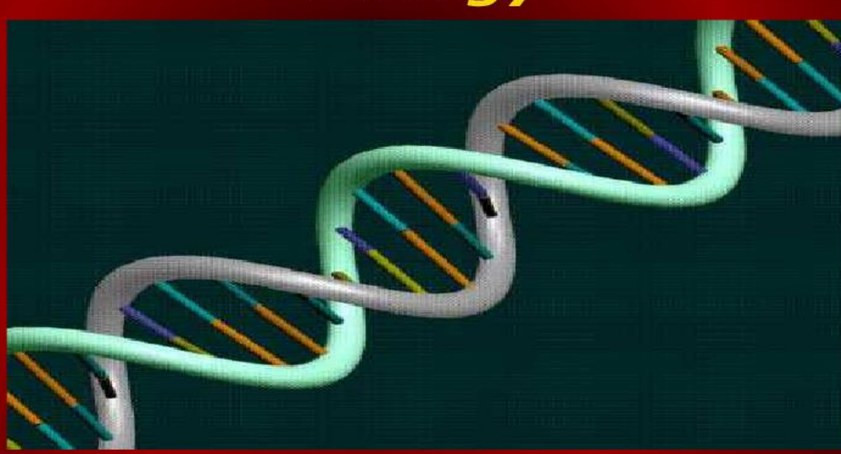


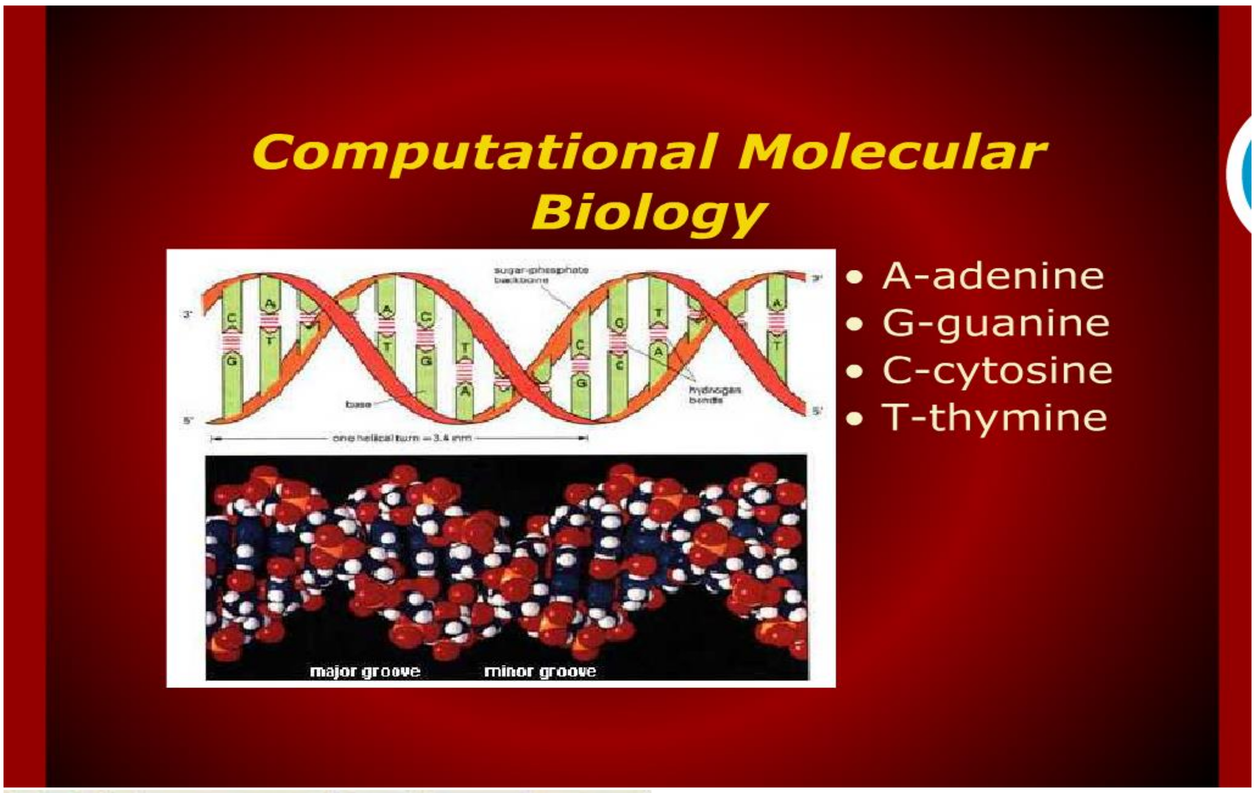


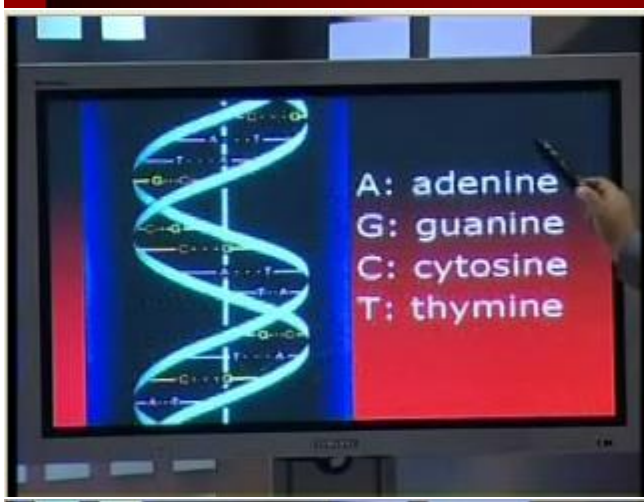
Edit Distance: Applications

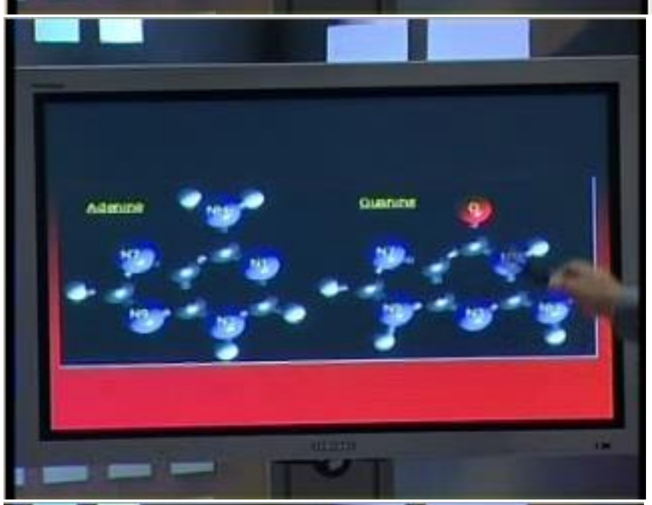
Plagiarism Detection
The edit distance provides an indication of similarity that might be too close in some situationsn.

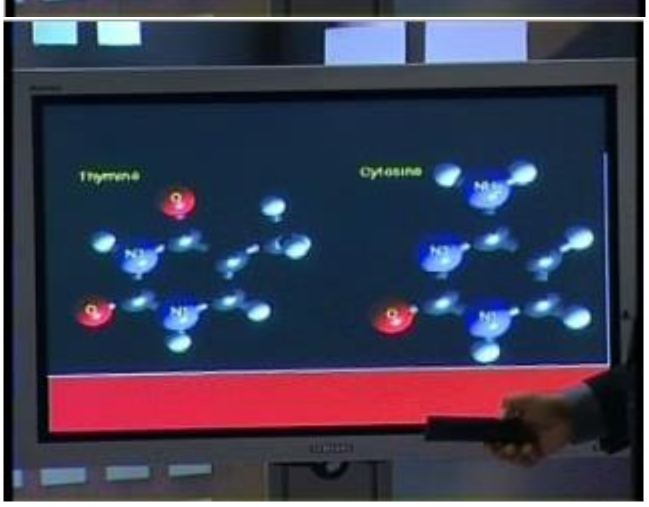
Computational Molecular Biology

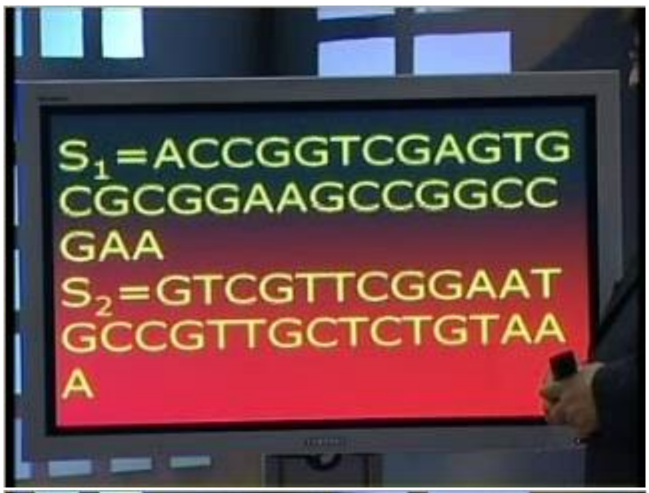



Computational Molecular Biology
sugar-phosphate backbone
hydrogen bonds
base
one helical turn = 3.4 nm
major groove
minor groove
• A-adenine
• G-guanine
• C-cytosine
• T-thymine

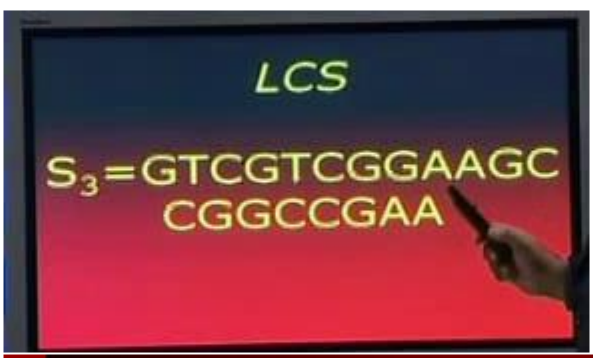

A: adenine
G: guanine
C: cytosine
T: thymine

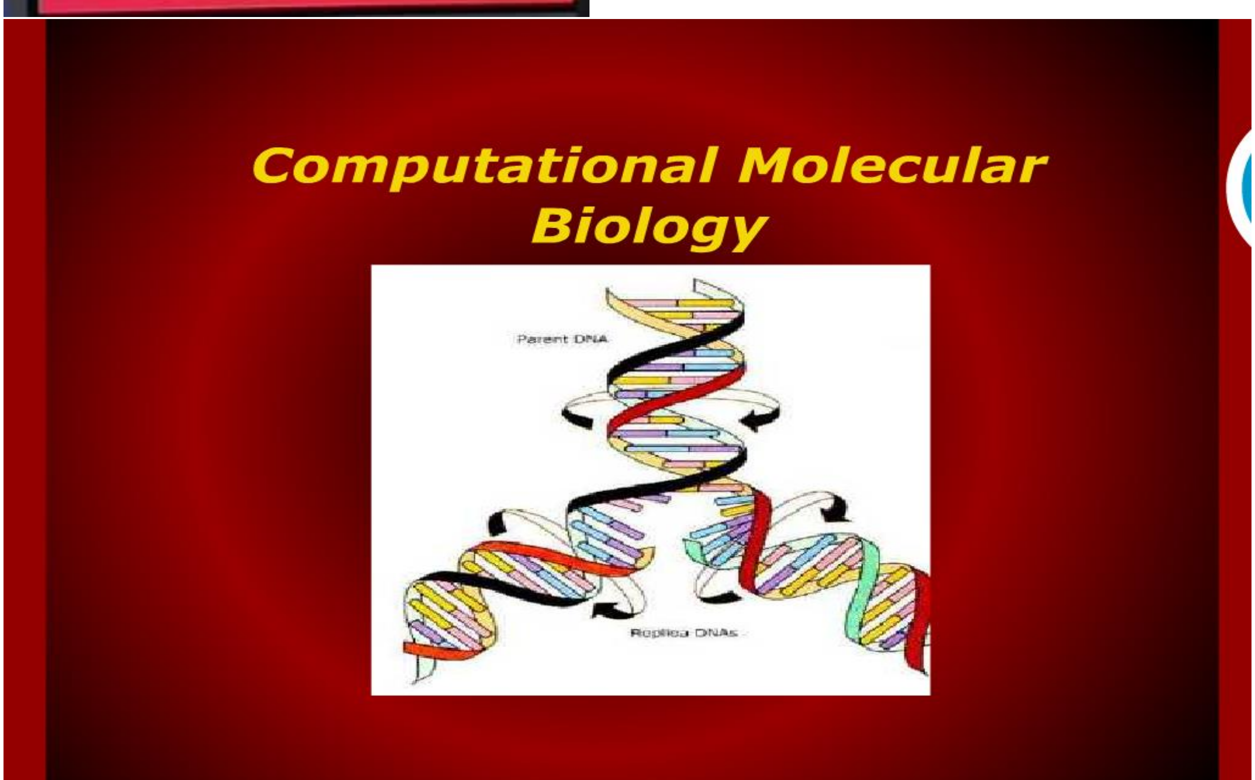

Adenine
Guanine


Thymine
Cytosine


S1=ACCGGTCGAGTG
CGCGGAAGCCGGCC
GAA
S2=GTCGTTCGGAAT
GCCGTTGCTCTGTAA
A


Human beta globin region on
chromosome 11 (73321)
gaattctaatctccctctcaaccctaca
gtcacccatttggtatattaaagatgtg
ttgtctactgtctagtatccctcaagta
gtgtcaggaattagtcatttaaatagtc
tgcaagccaggagtggtggctcatgtct
gtaattccagcactggagaggtagaagt
gggaggactgcttgagctcaagagtttg
atattatcctggacaacatagcaagacc


LCS
S3=GTCGTCGGAAGC
CGGCCGAA


Computational Molecular
Biology
Parent DNA
Replica DNAs

Edit Distance: Applications

Computational Molecular Biology
DNA Replication errors lead to nucleotide:

- Substitutions
- Insertions
- Deletion.

Edit Distance: Applications

Human beta globin region on chromosome 11

1	gaattctaat	ctccctctca	accctacagt	cacccatttg	gtatattaaa	gatgtgttgt
61	ctactgtcta	gtatccctca	agtagtgtca	ggaattagtc	atttaaatag	tctgcaagcc
121	aggagtggtg	gctcatgtct	gtaattccag	cactggagag	gtagaagtgg	gaggactgct
181	tgagctcaag	agtttgatat	tatcctggac	aacatagcaa	gacctcgtct	ctacttaaaa
241	aaaaaaaaat	tagccaggca	tgtgatgtac	acctgtagtc	ccagctactc	aggaggccga
301	aatgggagga	tcccttgagc	tcaggaggtc	aaggctgcag	tgagacatga	tcttgccact
361	gcactccagc	ctggacagca	gagtgaaacc	ttgcctcacg	aaacagaata	caaaaacaaa
421	caaacaaaaa	actgctccgc	aatgcgcttc	cttgatgctc	taccacatag	gtctgggtac
481	tttgtacaca	ttatctcatt	gctgttcgta	attgttagat	taattttgta	atattgatat
541	tattcctaga	aagctgaggc	ctcaagatga	taacttttat	tttctggact	tgtaatagct
601	ttctcttgta	ttcaccatgt	tgtaactttc	ttagagtagt	aacaatataa	agttattgtg
661	agtttttgca	aacacagcaa	acacaacgac	ccatatagac	attgatgtga	aattgtctat
721	tgtcaattta	tgggaaaaca	agtatgtact	ttttctacta	agccattgaa	acaggaataa
781	cagaacaaga	ttgaaagaat	acattttccg	aaattacttg	agtattatac	aaagacaagc

Edit Distance: Applications

841 acgtggacct gggaggaggg ttattgtcca tgactggtgt gtggagacaa atgcaggttt
901 ataatagatg ggatggcatc tagcgcaatg actttgccat cacttttaga gagctcttgg
961 ggaccccagt acacaagagg ggacgcaggg tatatgtaga catctcattc tttttcttag
1021 tgtgagaata agaatagcca tgacctgagt ttatagacaa tgagcccttt tctctctccc
1081 actcagcagc tatgagatgg cttgccctgc ctctctacta ggctgactca ctccaaggcc
1141 cagcaatggg cagggctctg tcagggcttt gatagcacta tctgcagagc cagggccgag
1201 aaggggtgga ctccagagac tctccctccc attcccgagc agggtttgct tatttatgca
1261 tttaaatgat atatttattt taaaagaaat aacaggagac tgcccagccc tggctgtgac
1321 atggaaacta tgtagaatat tttgggttcc attttttttt ccttctttca gttagaggaa
1381 aaggggctca ctgcacatac actagacaga aagtcaggag ctttgaatcc aagcctgatc
1441 atttccatgt catactgaga aagtccccac ccttctctga gcctcagttt ctctttttat
1501 aagtaggagt ctggagtaaa tgatttccaa tggctctcat ttcaatacaa aatttccgtt
1561 tattaaatgc atgagcttct gttactccaa gactgagaag gaaattgaac ctgagactca
1621 ttgactggca agatgtcccc agaggctctc attcagcaat aaaattctca ccttcaccca
1681 ggcccactga gtgtcagatt tgcatgcact agttcacgtg tgtaaaaagg aggatgcttc

Edit Distance: Applications

1741 tttcctttgt attctcacat acctttagga aagaacttag cacccttccc acacagccat
1801 cccaataact catttcagtg actcaaccct tgactttata aaagtcttgg gcagtataga
1861 gcagagatta agagtacaga tgctggagcc agaccacctg agtgattagt gactcagttt
1921 ctcttagtaa ttgtatgact cagtttcttc atctgtaaaa tggagggttt tttaattagt
1981 ttgtttttga gaaagggtct cactctgtca cccaaatggg agtgtagtgg caaaatctcg
2041 gctcactgca acttgcactt cccaggctca agcggtcctc ccacctcaac atcctgagta
2101 gctggaacca caggtacaca ccaccatacc tcgctaattt tttgtatttt tggtagagat
2161 ggggtttcac atgttacaca ggatggtctc agactccgga gctcaagcaa tctgcccacc
2221 tcagccttcc aaagtgctgg gattataagc atgattacag gagttttaac aggctcataa
2281 gattgttctg cagcccgagt gagttaatac atgcaaagag tttaaagcag tgacttataa
2341 atgctaacta ctctagaaat gtttgctagt attttttgtt taactgcaat cattcttgct
2401 gcaggtgaaa actagtgttc tgtactttat gcccattcat ctttaactgt aataataaaa
2461 ataactgaca tttattgaag gctatcagag actgtaatta gtgctttgca taattaatca
2521 tatttaatac tcttggattc tttcaggtag atactattat tatccccatt ttactacagt
2581 taaaaaaact acctctcaac ttgctcaagc atacactctc acacacacaa acataaacta
2641 ctagcaaata gtagaattga gatttggtcc taattatgtc tttgctcact atccaataaa
2701 tatttattga catgtacttc ttggcagtct gtatgctgga tgctggggat acaaagatgt

Edit Distance: Applications

72841	ggattaagaa	aatgtggcac	atatacacca	tggaatacta	tgcagccata	aaaaatgatg
72901	agttcatgtc	ctttgtaggg	acatggatga	agctggaaac	tatcattctc	agcaaactat
72961	cacaaggaca	ataaaccaaa	caccgcatgt	tctcactcat	aggtgggaat	tgaacaatga
73021	gaacacatgg	acacatgaag	aggaacatca	cactctgggg	actgttatgg	ggtggggggc
73081	aggggcaggg	atagcactag	gagatatacc	taatgctaaa	tgacgagtta	atgggtgcag
73141	cacaccaaca	tggcacatgt	atacatatat	aacaaacctg	ccgttgtgca	catgtaccct
73201	aaaacttgaa	gtataataat	aaaaaaaagt	tatcctatta	aaactgatct	cacacatccg
73261	tagagccatt	atcaagtctt	tctctttgaa	acagacagaa	atttagtgtt	ttctcagtca
73321	gttaac					

Edit Distance: Applications

Computational Molecular Biology

Similarities in DNA Sequences can provide

- Clue to common evolutionary origin
- Clue to common function

Edit Distance: Applications

Speech Recognition

- Algorithms similar to those for the edit-distance problem are used in some speech recognition systems.
- Find a close match between a new utterance and one in a library of classified utterances.

Edit Distance

A better way to display this editing process is to place the words above the other:

```
M A _ T H S
A _ R T _ S
```

Edit Distance

S	D	I	M	D	M
M	A	_	T	H	S
A	_	R	T	_	S

- The first word has a gap for every insertion (I) and the second word has a gap for every deletion (D).
- Columns with two different characters correspond to substitutions (S).

Edit Distance

S	D	I	M	D	M
M	A	_	T	H	S
A	_	R	T	_	S

- Edit transcript: A string over the alphabet M, S, I, D that describes a transformation of one string into another Example.

S	*D*	*I*	*M*	*D*	*M*
1+	1+	1+	0+	1+	0+ = 4