**Experiment No . 06: Internet Algorithm**

**DATE :- 12-05-2023**

**Aim:-** Write C program to implement the following using Internet Algorithm.

1. KMP Pattern Matching
2. BM Pattern Matching
3. Huffman Coding (Text Compression)
4. Longest common Subsequence. (Text Similarities)

**Theory**

* Internet algorithms are specifically designed algorithms used in various aspects of the internet to solve specific problems efficiently.
* These algorithms address challenges related to data transmission, routing, network optimization, information retrieval, and security, among others.
* Internet algorithms often consider factors such as latency, bandwidth, scalability, fault tolerance, and resource constraints to ensure efficient and reliable operations.

**Knuth-Morris-Pratt (KMP) Algorithm:**

* The KMP algorithm is used for pattern matching in strings efficiently.
* It avoids unnecessary comparisons by utilizing information from previous matches.
* It pre-processes the pattern to construct a lookup table (also called the "failure function") that helps skip unnecessary comparisons during matching.
* The algorithm compares the pattern with the input text character by character, utilizing the lookup table to determine the next position for comparison in case of a mismatch.
* The KMP algorithm achieves a time complexity of O(n + m), where n is the length of the input text and m is the length of the pattern, making it efficient for large-scale pattern matching.

**Boyer-Moore (BM) Algorithm:**

* The Boyer-Moore algorithm is another efficient pattern matching algorithm, particularly suitable for searching in large texts.
* It pre-processes the pattern and utilizes two heuristics: the "bad character rule" and the "good suffix rule."
* The bad character rule skips comparisons by shifting the pattern to align the rightmost occurrence of a mismatched character in the text with the corresponding position in the pattern.
* The good suffix rule utilizes information about matching suffixes in the pattern to skip unnecessary comparisons.
* The Boyer-Moore algorithm achieves a best-case time complexity of O(n/m) and an average-case time complexity of O(n + m), making it efficient for practical use.

**Huffman Coding:**

* Huffman coding is a compression algorithm used to encode data efficiently by assigning shorter codes to frequently occurring characters and longer codes to less frequent characters.
* The algorithm builds a binary tree (Huffman tree) based on the frequency of characters in the input data.
* The characters with higher frequency are assigned shorter codes, and those with lower frequency are assigned longer codes, ensuring prefix-free codes.
* Huffman coding achieves compression by representing the input data using the generated Huffman codes, reducing the overall number of bits required for storage or transmission.
* The time complexity of the Huffman coding algorithm is O(n log n), where n is the number of unique characters in the input.

**Longest Common Subsequence (LCS) Algorithm:**

* The LCS algorithm is used to find the longest common subsequence between two sequences, typically strings.
* It determines the longest subsequence that is present in both sequences but does not necessarily have to be contiguous.
* The algorithm utilizes dynamic programming to build a table that stores the lengths of the longest common subsequences for various subproblems.
* Starting from the end of the sequences, the algorithm fills the table by considering two cases: matching characters or non-matching characters.
* Finally, the algorithm traces back the table to reconstruct the LCS.
* The time complexity of the LCS algorithm is O(mn), where m and n are the lengths of the input sequences, making it efficient for practical use.

**A)KMP Pattern Matching**

**DATE :- 12-05-2023**

**Problem Statement**

Write a C program to implement KMP pattern matching for

T = **aabbbaababbbabab** and P = **bbaba**

**Algorithm**

**Algorithm KMPMatch(T,P)**:

Input: Strings T (text) with n characters and P (pattern) with m characters

Output: Starting index of the first substring of T matching P, or an indication

that P is not a substring of T

f ← KMPFailureFunction(P) // construct the failure function f for P

i ← 0

j ← 0

while i<n do

if P[j] = T[i] then

if j = m − 1 then

return i − m + 1 // a match!

i ← i + 1

j ← j + 1

else if j > 0 // no match, but we have advanced in P then

j ← f(j − 1) // j indexes just after prefix of P that must match

else

i ← i + 1

return “There is no substring of T matching P.”

**Algorithm KMPFailureFunction(P):**

Input: String P (pattern) with m characters

Output: The failure function f for P, which maps j to the length of the longest

prefix of P that is a suffix of P[1..j]

i ← 1

j ← 0

f(0) ← 0

while i<m do

if P[j] = P[i] then

// we have matched j + 1 characters

f(i) ← j + 1

i ← i + 1

j ← j + 1

else if j > 0 then

// j indexes just after a prefix of P that must match

j ← f(j − 1)

else

// we have no match here

f(i) ← 0

i ← i + 1

**Time and Space Complexity**

Time Complexity = **O(m+n)**

Space complexity = **O(m)**

Where , n is the length of text and m is the length of pattern .

**Program:**

#include <stdio.h>

#include <string.h>

#define MAX 100

int f[MAX];

void KMPFailureFunction(char P[])

{

    int i = 1, j = 0, m = strlen(P);

    f[0] = 0;

    while (i < m)

    {

        if (P[j] == P[i])

        {

            f[i] = j + 1;

            i++;

            j++;

        }

        else if (j > 0)

        {

            j = f[j - 1];

        }

        else

        {

            f[i] = 0;

            i++;

        }

    }

}

int KMPMatch(char T[], char P[])

{

    int n = strlen(T), m = strlen(P);

    int i = 0, j = 0;

    KMPFailureFunction(P);

    while (i < n)

    {

        if (P[j] == T[i])

        {

            if (j == m - 1)

            {

                return i - m + 1;

            }

            i++;

            j++;

        }

        else if (j > 0)

        {

            j = f[j - 1];

        }

        else

        {

            i++;

        }

    }

    return -1;

}

int main()

{

    int index;

    char T[MAX], P[MAX];

    printf("Enter the string T:");

    scanf("%s", &T);

    printf("Enter the string P:");

    scanf("%s", &P);

    index = KMPMatch(T, P);

    printf("\n");

    if (index == -1)

    {

        printf("There is no substring of %s matching %s\n", T, P);

    }

    else

    {

        printf("Pattern %s is found at index %d in String %s", P, index + 1, T);

    }

    return 0;

}

**Output:**

c:\P Jeevesh Naidu\college\second year\lV sem\madf codes>cd "c:\P Jeevesh Naidu\college\second year\lV sem\madf codes\" && gcc kmp.c -o kmp && "c:\P Jeevesh Naidu\college\second year\lV sem\madf codes\"kmp

Enter the string T:aabbbaababbbabab

Enter the string P:bbaba

Pattern bbaba is found at index 11 in String aabbbaababbbabab

**B)** **Boyer Moore pattern Matching**

**DATE :- 15-05-2023**

**Problem Statement**

Write a C program to implement Boyer Moore pattern matching for

T = 1123114234112113 and P = 4112113

**Algorithm**

**Algorithm BMMatch(T,P):**

Input: Strings T (text) with n characters and P (pattern) with m characters

Output: Starting index of the first substring of T matching P, or an indication

that P is not a substring of T

compute function last

i ← m − 1

j ← m − 1

repeat

if P[j] = T[i] then

if j = 0 then

return i // a match!

else

i ← i − 1

j ← j − 1

else

i ← i + m − min(j, 1 + last(T[i])) // jump step

j ← m − 1

until i>n − 1

return “There is no substring of T matching P.”

**Time and Space Complexity**

Time Complexity = **O(m+n)**

Space complexity = **O(m+k)**

Where , n is the length of text and m is the length of pattern .

**Program:**

#include <stdio.h>

#include <string.h>

#define MAX 256

int lastarr[MAX];

int min(int a, int b)

{

    return a < b ? a : b;

}

void last(char pattern[MAX], int m)

{

    // Index of the last (right most) occurrence of c in pattern.

    // Otherwise we define last(c) = -1.

    for (int i = 0; i < MAX; i++)

        lastarr[i] = -1;

    for (int i = 0; i < m; i++)

        lastarr[(int)pattern[i]] = i;

}

void BM(char test[], char pattern[], int n, int m)

{

    last(pattern, m);

    int i, j;

    i = m - 1;

    j = m - 1;

    do

    {

        if (pattern[j] == test[i])

        {

            printf("%d of test matches with %d of pattern.\n", i, j);

            if (!j)

            {

                printf("Pattern %s found at index %d in string %s.\n", pattern, i, test);

                return;

            }

            else

            {

                i--;

                j--;

            }

        }

        else

        {

            printf("%d of test mismatches with %d of pattern.\n", i, j);

            i = i + m - min(j, 1 + lastarr[(int)test[i]]);

            j = m - 1;

        }

    } while (i <= n - 1);

    printf("Pattern %s not present in string %s!\n", pattern, test);

}

int main()

{

    char test[50], pattern[50];

    printf("Enter test string - ");

    scanf("%s", test);

    printf("Enter pattern string - ");

    scanf("%s", pattern);

    BM(test, pattern, strlen(test), strlen(pattern));

    return 0;

}

**Output:**

c:\P Jeevesh Naidu\college\second year\lV sem\madf codes>cd "c:\P Jeevesh Naidu\college\second year\lV sem\madf codes\" && gcc bm.c -o bm && "c:\P Jeevesh Naidu\college\second year\lV sem\madf codes\"bm

Enter test string - 1123114234112113

Enter pattern string - 4112113

6 of test mismatches with 6 of pattern.

12 of test mismatches with 6 of pattern.

15 of test matches with 6 of pattern.

14 of test matches with 5 of pattern.

13 of test matches with 4 of pattern.

12 of test matches with 3 of pattern.

11 of test matches with 2 of pattern.

10 of test matches with 1 of pattern.

9 of test matches with 0 of pattern.

Pattern 4112113 found at index 9 in string 1123114234112113.

1. **Huffman Coding (Text Compression)**

**DATE :- 19-05-2023**

**Problem Statement**

Write a C program to implement Huffman Encoding for

“j is the position of the ppartitioning element”.

**Algorithm**

**Algorithm Huffman(C):**

Input: A set, C, of d characters, each with a given weight, f(c)

Output: A coding tree, T, for C, with minimum total path weight

Initialize a priority queue Q.

for each character c in C do

Create a single-node binary tree T storing c.

Insert T into Q with key f(c).

while Q.size() > 1 do

f1 ← Q.minKey()

T1 ← Q.removeMin()

f2 ← Q.minKey()

T2 ← Q.removeMin()

Create a new binary tree T with left subtree T1 and right subtree T2.

Insert T into Q with key f1 + f2.

return tree Q.removeMin()

**Time and Space Complexity**

Time Complexity = **O(nlogn+m)**

Space complexity = **O(n+mlogn)**

**Program:**

#include <stdio.h>

#include <stdlib.h>

#include <string.h>

#define sint(x) scanf("%d", &x)

#define N 100

struct node

{

    int freq;

    char c;

    struct node \*next, \*left, \*right;

};

struct node \*start = NULL;

greatest = 0; // this holds the greatest freq

void addatend(char c[], int f[], int n)

{

    for (int i = 0; i < n; i++)

    {

        if (start == NULL)

        {

            start = (struct node \*)malloc(sizeof(struct node));

            start->c = c[i], start->freq = f[i], start->next = NULL, start->left = NULL, start->right = NULL;

        }

        else

        {

            struct node \*ptr = start;

            while (ptr->next != NULL)

                ptr = ptr->next;

            struct node \*temp = (struct node \*)malloc(sizeof(struct node));

            temp->c = c[i], temp->freq = f[i];

            temp->next = temp->left = temp->right = NULL, ptr->next = temp;

        }

    }

}

void showlist()

{

    struct node \*ptr = start;

    while (ptr != NULL)

    {

        printf("(%c, %d), ", ptr->c, ptr->freq);

        ptr = ptr->next;

    }

    printf("\n");

}

struct node \*delfront()

{

    struct node \*temp = start;

    start = start->next;

    return temp;

}

void addafterliketerms(struct node \*temp)

{

    // check if there is only one

    if (start == NULL)

        start = temp;

    else if (start->next == NULL) // only one term

    {

        if (temp->freq < start->freq)

        {

            temp->next = start;

            start = temp;

        }

        else

        {

            temp->next = start->next;

            start->next = temp;

        }

    }

    else

    {

        int found = 0;

        struct node \*at, \*ptr = start;

        while (ptr != NULL)

        {

            if (ptr->freq == temp->freq)

            {

                found = 1, at = ptr;

                break;

            }

            ptr = ptr->next;

        }

        ptr = start;

        if (found)

        {

            struct node \*ptr = at;

            while (ptr->next->freq == temp->freq && ptr->next != NULL)

                ptr = ptr->next;

            temp->next = ptr->next;

            ptr->next = temp;

        }

        else

        {

            ptr = start;

            while (ptr->next->freq <= temp->freq && ptr != NULL)

                ptr = ptr->next;

            if (ptr != NULL)

            {

                temp->next = ptr->next;

                ptr->next = temp;

            }

            else

            {

                ptr = start;

                while (ptr->next != NULL)

                    ptr = ptr->next;

                temp->next = ptr->next;

                ptr->next = temp;

            }

        }

    }

}

void ins\_end(struct node \*temp)

{

    if (start == NULL)

        temp->next = NULL, start = temp;

    else

    {

        struct node \*ptr = start;

        while (ptr->next != NULL)

            ptr = ptr->next;

        temp->next = ptr->next;

        ptr->next = temp;

    }

}

void huffman()

{

    if (!start)

        return;

    while (start->next != NULL)

    {

        struct node \*a = delfront(), \*b = delfront(), \*temp = (struct node \*)malloc(sizeof(struct node));

        temp->c = '#', temp->freq = a->freq + b->freq, temp->left = a, temp->right = b, temp->next = NULL;

        if (temp->freq >= greatest)

            greatest = temp->freq, ins\_end(temp);

        else

            addafterliketerms(temp);

        showlist();

        printf("\n");

    }

}

void printcodes(struct node \*root, int path[], int stringlen)

{

    if (root == NULL)

        return;

    if (root->left == NULL && root->right == NULL)

    {

        for (int i = 0; i < stringlen; i++)

            printf("%d", path[i]);

        printf("\t%c\n", root->c);

    }

    // left tree

    path[stringlen] = 0;

    printcodes(root->left, path, stringlen + 1);

    // right tree

    path[stringlen] = 1;

    printcodes(root->right, path, stringlen + 1);

}

void findgreatest()

{

    struct node \*ptr = start;

    while (ptr->next != NULL)

        ptr = ptr->next;

    greatest = ptr->freq;

}

int main()

{

    // char arr[] = {'j', 'f', 'a', 'r', 'g', 'g', 'm', 's', 'h', 'p', 'o','n', 'e', 'i', 't', 'sp'};

    // int freq[] = {1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 4, 4, 5, 6, 6, 7};

    // j is the position of the partitioning element

    char arr[17] = {'j', 'f', 'a', 'r', 'g', 'l', 'm', 's', 'h', 'p', 'o','n', 'e', 'i', 't', ' '};

    int freq[17] = {1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 4, 4, 5, 6, 6, 7};

    printf("%d\n", strlen(arr));

    addatend(arr, freq, strlen(arr));

    showlist();

    findgreatest();

    huffman();

    int path[N];

    printf("Huffman Codes\n");

    printcodes(start, path, 0);

    return 0;

}

**Output:**

c:\P Jeevesh Naidu\college\second year\lV sem\madf codes>cd "c:\P Jeevesh Naidu\college\second year\lV sem\madf codes\" && gcc tempCodeRunnerFile.c -o tempCodeRunnerFile && "c:\P Jeevesh Naidu\college\second year\lV sem\madf codes\"tempCodeRunnerFile

(j, 1), (f, 1), (a, 1), (r, 1), (g, 1), (l, 1), (m, 1), (s, 2), (h, 2), (p, 2), (o, 4), (n, 4), (e, 5), (i, 6), (t, 6), ( , 7),

(a, 1), (r, 1), (g, 1), (l, 1), (m, 1), (s, 2), (h, 2), (p, 2), (#, 2), (o, 4), (n, 4), (e, 5), (i, 6), (t, 6), ( , 7),

(g, 1), (l, 1), (m, 1), (s, 2), (h, 2), (p, 2), (#, 2), (#, 2), (o, 4), (n, 4), (e, 5), (i, 6), (t, 6), ( , 7)

(m, 1), (s, 2), (h, 2), (p, 2), (#, 2), (#, 2), (#, 2), (o, 4), (n, 4), (e, 5), (i, 6), (t, 6), ( , 7),

(h, 2), (p, 2), (#, 2), (#, 2), (#, 2), (#, 3), (o, 4), (n, 4), (e, 5), (i, 6), (t, 6), ( , 7),

(#, 2), (#, 2), (#, 2), (#, 3), (o, 4), (n, 4), (#, 4), (e, 5), (i, 6), (t, 6), ( , 7),

(#, 2), (#, 3), (o, 4), (n, 4), (#, 4), (#, 4), (e, 5), (i, 6), (t, 6), ( , 7),

(o, 4), (n, 4), (#, 4), (#, 4), (e, 5), (#, 5), (i, 6), (t, 6), ( , 7),

(#, 4), (#, 4), (e, 5), (#, 5), (i, 6), (t, 6), ( , 7), (#, 8),

(e, 5), (#, 5), (i, 6), (t, 6), ( , 7), (#, 8), (#, 8),

(i, 6), (t, 6), ( , 7), (#, 8), (#, 8), (#, 10),

( , 7), (#, 8), (#, 8), (#, 10), (#, 12),

(#, 8), (#, 10), (#, 12), (#, 15),

(#, 12), (#, 15), (#, 18),

(#, 18), (#, 27),

(#, 45),

Huffman Codes

0000 h

0001 p

00100 j

00101 f

00110 a

00111 r

010 e

01100 g

01101 l

01110 m

01111 s

100 i

101 t

110

1110 o

1111 n

**D)LCS (Text Similarity)**

**DATE :- 22-05-2023**

**Problem Statement**

Write a C program to implement LCS for X = KLOKMKNKLOK and Y = KLLKNKLLKNYY

**Algorithm**

**Algorithm LCS(X, Y ):**

Input: Strings X and Y with n and m elements, respectively

Output: For i = 0,...,n − 1, j = 0,...,m − 1, the length L[i, j] of a longest

common subsequence of X[0..i] and Y [0..j]

for i ← −1 to n − 1 do

L[i, −1] ← 0

for j ← 0 to m − 1 do

L[−1, j] ← 0

for i ← 0 to n − 1 do

for j ← 0 to m − 1 do

if X[i] = Y [j] then

L[i, j] ← L[i − 1, j − 1] + 1

else

L[i, j] ← max{L[i − 1, j] , L[i, j − 1]}

return array L

**Time and Space Complexity**

Time Complexity = **O(mn)**

Space complexity = **O(mn)**

Where , n and m are the lengths of the strings.

**Program:**

#include <stdio.h>

#include <string.h>

#define N 256

int L[N][N];

int smax(int a, int b)

{

    return a > b ? a : b;

}

void lcs(char x[], char y[], int n, int m)

{

    for (int i = 0; i <= m; i++)

        L[i][0] = 0;

    for (int j = 0; j <= n; j++)

        L[0][j] = 0;

    for (int i = 1; i <= m; i++)

    {

        for (int j = 1; j <= n; j++)

        {

            if (x[j - 1] == y[i - 1])

                L[i][j] = L[i - 1][j - 1] + 1;

            else

                L[i][j] = smax(L[i - 1][j], L[i][j - 1]);

        }

    }

}

void sequence(int ss[], int m, int n)

{

    int i = m, j = n, k = m;

    while (i > 0)

    {

        if (L[i][j] == L[i - 1][j])

            ss[k] = 0, k--, i--;

        else

            ss[k] = 1, i--, j--, k--;

    }

}

void show(char y[], int ss[], int m)

{

    for (int i = 1, j = 0; i <= m, j < m; i++, j++)

    {

        if (ss[i])

            printf("%c", y[j]);

    }

    printf("\n");

}

int main()

{

    char x[] = "KLOKMKNKLOK", y[] = "KLLKNKLLKNYY";

    int m = strlen(y), n = strlen(x), ss[strlen(y)];

    lcs(x, y, n, m), sequence(ss, m, n);

    printf("Length of the longest common subsequence is: %d\nThe longest subsequence is: ", L[m][n]), show(y, ss, m);

    return 0;

}

**Output:**

c:\P Jeevesh Naidu\college\second year\lV sem\madf codes>cd "c:\P Jeevesh Naidu\college\second year\lV sem\madf codes\" && gcc lcs.c -o lcs && "c:\P Jeevesh Naidu\college\second year\lV sem\madf codes\"lcs

Length of the longest common subsequence is: 7

The longest subsequence is: KLKNKLK

**Conclusion:**

Internet algorithms were studied . The program for

1. KMP Pattern Matching
2. BM Pattern Matching
3. Huffman Coding
4. Longest Common Subsequence

were studied and implemented successfully.