

WEB MINING

Soumok Dutta
16BCE1184

HITS ALGORITHM

Code:

```
import numpy as np
def authority_hub_score(outlinks):
    size = outlinks.shape[0]
    hub_scores = [1.0 for i in range(size)]
    authority_scores = [1.0 for i in range(size)]
    print(hub_scores)

    for _ in range(100):
        for j in range(size):
            temp_auth = 0.0
            for i in range(size):
                if outlinks[i][j] == 1:
                    temp_auth += hub_scores[i]
            authority_scores[j] = temp_auth

        auth_sum = sum(authority_scores)

        for i in range(len(authority_scores)):
            authority_scores[i] /= auth_sum

        for i in range(size):
            temp_hub = 0.0
            for j in range(size):
                if outlinks[i][j] == 1:
                    temp_hub += authority_scores[j]
            hub_scores[i] = temp_hub

        hub_sum = sum(hub_scores)
        for i in range(len(hub_scores)):
            hub_scores[i] /= hub_sum

    return authority_scores, hub_scores

n = int(input('Enter the size of the matrix:\t'))
outlinks = []
```

```

for i in range(n*n):
    temp = int(input('Enter the element:\t'))
    outlinks.append(temp)
outlinks = np.reshape(outlinks, (n, n))
authority_scores, hub_scores = authority_hub_score(outlinks)
print("Authority Scores:")
for i in (authority_scores):
    print(round(i, 4))
print("Hub Scores:")
for i in (hub_scores):
    print(round(i, 4))

```

PAGERANK ALGORITHM

Code:

```

import numpy as np

def calculate_PageRank(outlinks):

    d = 0.85
    size = outlinks.shape[0]
    page_ranks = [1 for i in range(size)]
    out_degrees = []
    for i in range(size):
        sums = 0
        for j in range(size):
            sums += outlinks[i][j]
        out_degrees.append(sums)
    print('Initial page ranks:')
    print(page_ranks)

    for _ in range(100):
        for j in range(size):
            temp = 0
            for i in range(size):
                if outlinks[i][j] == 1:
                    temp += page_ranks[i] / out_degrees[i]
            temp *= d
            temp += (1-d)
            page_ranks[j] = round(temp, 4)

    return page_ranks

outlinks = [1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0]
outlinks = np.reshape(outlinks, (5, 5))
page_ranks = calculate_PageRank(outlinks)
print()

```

```
print("The converged page rank is:")
print(page_ranks)
print()
sums = 0
for i in page_ranks:
    sums += i
print("The sum of page ranks is: ', round(sums, 4))
```