movies.dat

1|Toy Story (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Toy%20Story%20(1995)|0|0|0|1|1|1|0|0|0|0|0|0|0|0|0|0|0|0|0

2|GoldenEye (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?GoldenEye%20(1995)|0|1|1|0|0|0|0|0|0|0|0|0|0|0|0|0|1|0|0

3|Four Rooms (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Four%20Rooms%20(1995)|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|1|0|0

4|Get Shorty (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Get%20Shorty%20(1995)|0|1|0|0|0|1|0|0|1|0|0|0|0|0|0|0|0|0|0

5|Copycat (1995)|01-Jan-1995||http://us.imdb.com/M/title-exact?Copycat%20(1995)|0|0|0|0|0|0|1|0|1|0|0|0|0|0|0|0|1|0|0

test.dat

1 2 3 100

1 4 3 100

2 1 2 100

2 2 3 100

2 3 2 100

2 4 4 100

3 2 4 100

3 3 2 100

3 4 5 100

4 3 4 100

4 4 3 100

4 5 4 100

5 1 3 100

5 2 3 100

5 4 5 100

5 5 5 100

result

List of movies are:

{ '1': 'Toy Story (1995)',

'2': 'GoldenEye (1995)',

'3': 'Four Rooms (1995)',

'4': 'Get Shorty (1995)',

'5': 'Copycat (1995)'}

The preference list is:

{ '1': {'GoldenEye (1995)': 3.0,

'Get Shorty (1995)': 3.0},

'2': { 'Four Rooms (1995)': 2.0,

'Get Shorty (1995)': 4.0,

'GoldenEye (1995)': 3.0,

'Toy Story (1995)': 2.0},

'3': { 'Four Rooms (1995)': 2.0,

'Get Shorty (1995)': 5.0,

'GoldenEye (1995)': 4.0},

'4': { 'Copycat (1995)': 4.0,

'Four Rooms (1995)': 4.0,

'Get Shorty (1995)': 3.0},

'5': { 'Copycat (1995)': 5.0,

'Get Shorty (1995)': 5.0,

'GoldenEye (1995)': 3.0,

'Toy Story (1995)': 3.0}}

==============================================

Compute pair-wise similarity, similarity between users "1" and "2"

similarity between user 1 and user 2

prefs["1"]: {'GoldenEye (1995)': 3.0,

'Get Shorty (1995)': 3.0}

prefs["2"]: { 'Four Rooms (1995)': 2.0,

'Get Shorty (1995)': 4.0,

'GoldenEye (1995)': 3.0,

'Toy Story (1995)': 2.0}

Shared items, si: ['GoldenEye (1995)', 'Get Shorty (1995)']

distance = sqrt( (prefs["1"][si[0]]-prefs["2"][si[0]])^2 +

(prefs["2"][si[1]]-prefs["2"][si[1]]^2))

= sqrt( (prefs["1"]['GoldenEye'] - prefs["2"]['GoldenEye'])^2 +

(prefs["2"]['Get Shorty'] - prefs["2"]['Get Shorty']^2))

= sqrt ((3.0-3.0)^2 + (3.0-4.0)^2) = 1.0

similarity = 1/(1+distance) = 0.5

==============================================

Final rating:

[(4.25, 'Copycat (1995)'),

(3.1055728090000843, 'Four Rooms (1995)'),

(2.4000000000000004, 'Toy Story (1995)')]