

Epydoc

API Documentation

March 4, 2016

Contents

Contents	1
1 Module nonNumericalRecordPairsMatching	2
1.1 Variables	2
1.2 Class fieldPairSim	2
1.2.1 Methods	2
1.3 Class nonNumericalRecordPairsMatching	2
1.3.1 Methods	2
1.3.2 Properties	3
1.3.3 Class Variables	3

1 Module nonNumericalRecordPairsMatching

Created on Wed Dec 29 10:45:23 2015

Author: fubao

1.1 Variables

Name	Description
Intermediate_DirFiles	Value: ['intermediateOutput/nonNumericalInterOutput/', 'intermed...']
__package__	Value: None

1.2 Class fieldPairSim

1.2.1 Methods

```
__init__(self, fieldA, fieldB, value)
```

1.3 Class nonNumericalRecordPairsMatching

```
object └─ nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching
```

1.3.1 Methods

```
__init__(self)
```

x.__init__(...) initializes x; see help(type(x)) for signature

Overrides: object.__init__ extit(inherited documentation)

```
generatePairsTupleTobeMatched(self, tbFieldAllNonNumericalValuesMap)
```

```
cosinSimiAmongRecord(self, recA, recB, vABCacheMap)
```

```
getParsedSetMap(self, lsValA, lsValB, prefixLength)
```

```
filterColumnsWithSample(self, samplesFlag, prA, prB, lsValA, lsValB, prefixLength,
sampleNumofRecords, threshold, pairsNameLstA, pairsNameLstB, cosResLst, cachedSet)
```

```
getSamplesNonnumericalCosSimiRecordWise(self, pair, samplesFlag,
tbFieldAllNonNumericalValuesMap, prefixLength, sampleRecordsNum,
recordPrSimiThreshold, finalNonNumericalOutputDir)
```

```
getSamplesNonnumericalCosSimiRecordWiseMap(self, pair)
```

```
readSamplesResultTopKMatchingRatio(self, outFileNonNumericalRatioScoreSample,
outFileNonNumericalRatioScoreAll)
```

```
getAllNonnumericalCosSimiRecordWiseMap(self, pair)
```

```
sortAndWritetoFile(self, lstTopPairsTobeAllMatched, outFile)
```

```
multithreadgetAllNonNumericalCosinSimi(self, pairsAllTop,
outFileNonNumericalRatioScoreAll)
```

```
multithreadgetNonNumericalCosinSimi(self, pairsTupleTobeMatched,
outFileNonNumericalRatioScoreAll)
```

Inherited from object

```
__delattr__(), __format__(), __getattr__(), __hash__(), __new__(), __reduce__(), __reduce_ex__(),
__repr__(), __setattr__(), __sizeof__(), __str__(), __subclasshook__()
```

1.3.2 Properties

Name	Description
<i>Inherited from object</i>	
__class__	

1.3.3 Class Variables

Name	Description
tbFieldAllNonNumericalValuesMap	Value: {}
prefixLength	Value: 2
sampleRecordsNum	Value: 4000
recordPrSimiThreshold	Value: 0.428
finalNonNumericalOutputDir	Value: ''
threadNum	Value: 4

Index

nonNumericalRecordPairsMatching (*module*),
2–3

nonNumericalRecordPairsMatching.fieldPairSim
(*class*), 2

nonNumericalRecordPairsMatching.fieldPairSim.__init__
(*method*), 2

nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching
(*class*), 2–3

nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching.cosinSimiAmongRecord
(*method*), 2

nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching.filterColumnsWithSample
(*method*), 2

nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching.generatePairsTupleTobeMat
(*method*), 2

nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching.getAllNonnumericalCosSimiF
(*method*), 3

nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching.getParsedSetMap
(*method*), 2

nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching.getSamplesNonnumericalCos
(*method*), 2

nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching.getSamplesNonnumericalCos
(*method*), 3

nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching.multithreadgetAllNonNume
(*method*), 3

nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching.multithreadgetNonNumerical
(*method*), 3

nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching.readSamplesResultTopKMa
(*method*), 3

nonNumericalRecordPairsMatching.nonNumericalRecordPairsMatching.sortAndWritetoFile
(*method*), 3