Precision-Recall Analysis (norm-P-score)

March 14, 2017

Created by Alberto Ueda on 2017-01-30

0.1 Methods

```
In [1]: %pylab inline
        import pandas as pd
        import numpy as np

        import matplotlib.pyplot as plt

Populating the interactive namespace from numpy and matplotlib
```

```
In [265]: def plot_prec_recall(df, name):

              axes = df.plot(label='H-index', x='Recall.0@k',y='Precision.0@k', style='r-', grid
              df.plot(label='P-score', x='Recall.1@k', y='Precision.1@k', style='g-', grid=3, ax
              df.plot(label='norm-P-score', x='Recall.2@k', y='Precision.2@k', style='b-', grid=

              axes.set_ylim(0, 1.1)
              axes.set_xlim(0, 1.05)
              axes.set_xlabel("Recall@k")
              axes.set_title(name)
```

```
In [108]: def compute_prec_recall(rankings, vsubareas, subarea, dirname):

              result = rankings.merge(
                  vsubareas, on='Vkey', how='left').merge(
                  vsubareas, left_on='Vkey.1', right_on='Vkey', how='left').merge(
                  vsubareas, left_on='Vkey.2', right_on='Vkey', how='left')

              result.rename(columns={'Rank':'Rank.0', 'Vkey_x':'Vkey.0', 'Subareas_x':'Subareas.
                                     'Subareas_y':'Subareas.1', 'Subareas':'Subareas.2'}

              totalrel = float(vsubareas[vsubareas['Subareas'].str.contains(subarea)].shape[0])

              for i in range(0, 3):
                  result[subarea + str(i)] = result['Subareas.'+ str(i)].apply(lambda x: subarea
```

1

```python
            foundrel = 0

            for index, row in result.iterrows():
                if row[subarea + str(i)] == True:
                    foundrel += 1
                result.loc[index, 'Precision.' + str(i) + '@k'] = foundrel / float(index+1
                result.loc[index, 'Recall.' + str(i) + '@k'] = foundrel / totalrel

            del result[subarea + str(i)]

        result = result[[
                'Rank.0','Vkey.0','Subareas.0', 'Precision.0@k', 'Recall.0@k',
                'Rank.1','Vkey.1','Subareas.1', 'Precision.1@k', 'Recall.1@k',
                'Rank.2','Vkey.2','Subareas.2', 'Precision.2@k', 'Recall.2@k',
        ]]

        result.to_csv('expert-subareas-classification/' + dirname + '/' + subarea + '_prec
        plot_prec_recall(result, subarea)

        return result
```

```python
In [237]: def show_me(subarea, dirname, min_votes):
              vs = pd.read_csv('expert-subareas-classification/' + dirname + '/venues_subareas-'
              rankings = pd.read_csv('input/march-9-the-right-50/' + subarea + '.csv',
                                     usecols = ['Rank', 'Vkey', 'Rank.1', 'Vkey.1', 'Rank.2',

              compute_prec_recall(rankings, vs, subarea, dirname)
```

0.2 Plotting I

```python
In [155]: show_me('Information retrieval', 'ir-db-cn-berthier', 2)
```

Information retrieval
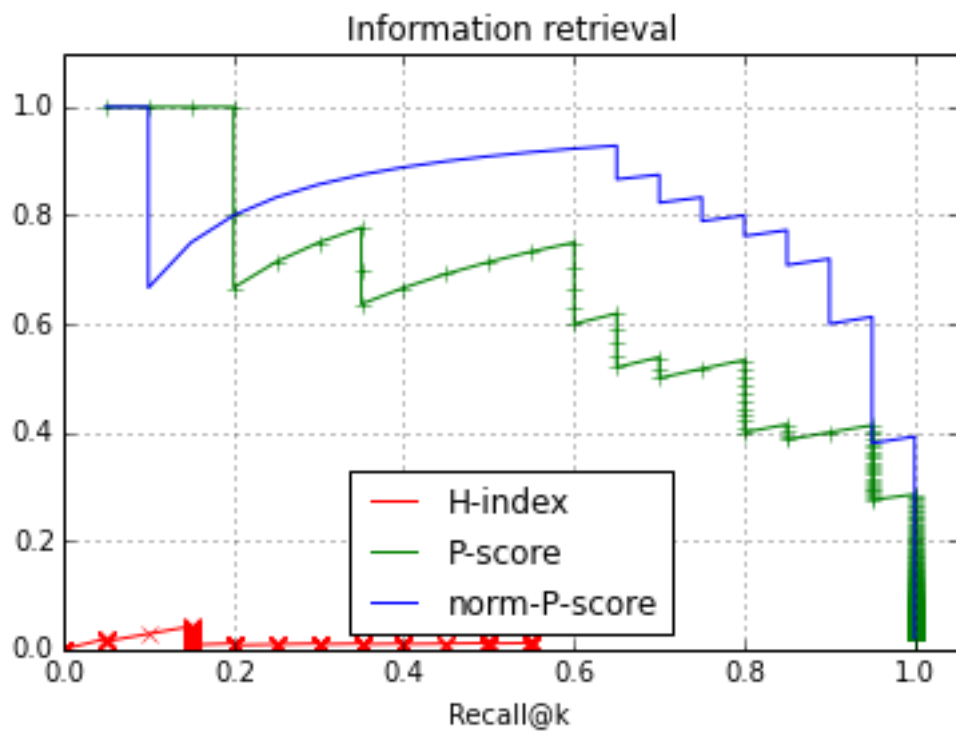
In [109]: show_me('Database', 'ir-db-cn-berthier', 2)
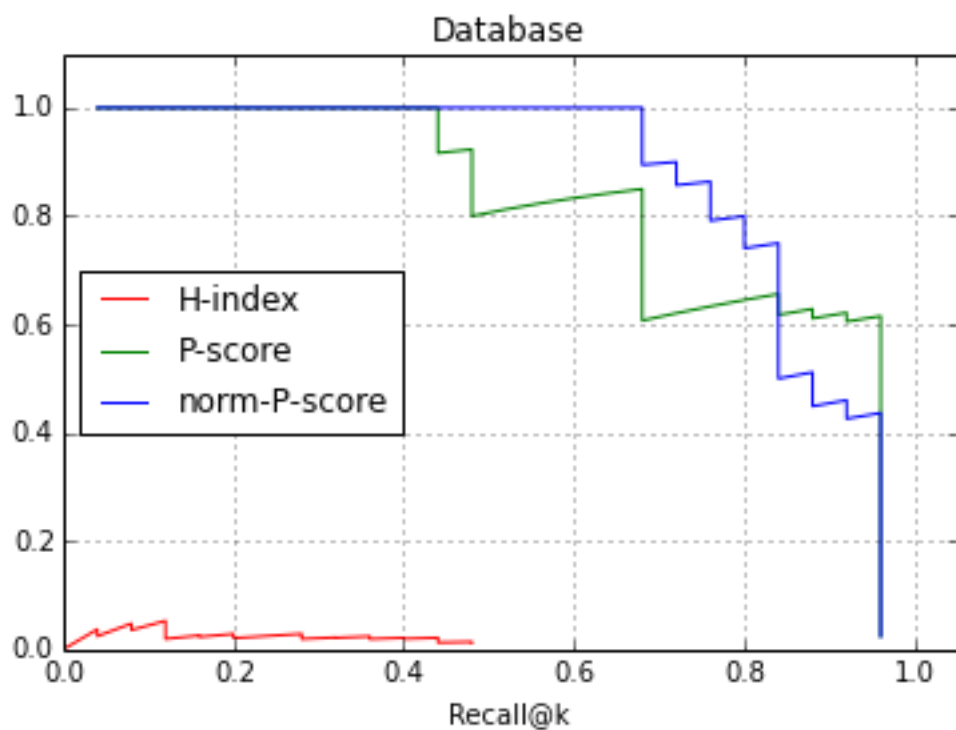
Database

```
In [111]: show_me('Computer network', 'ir-db-cn-berthier', 2)
```


0.3 Plotting II
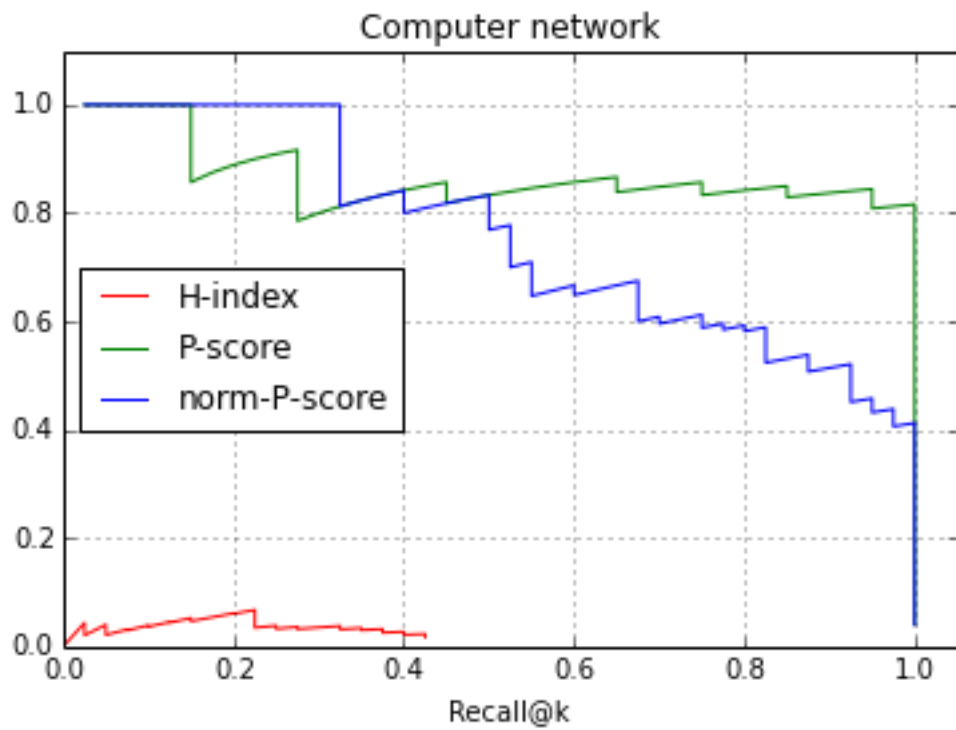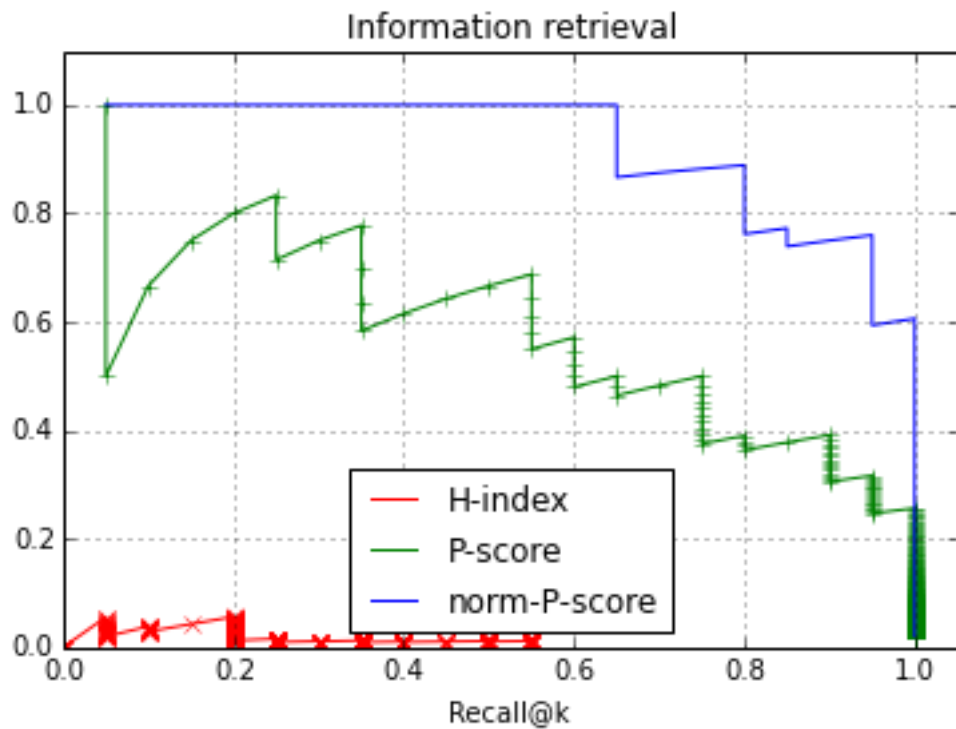
```
In [159]: show_me('Information retrieval', 'ir-students', 2)
```
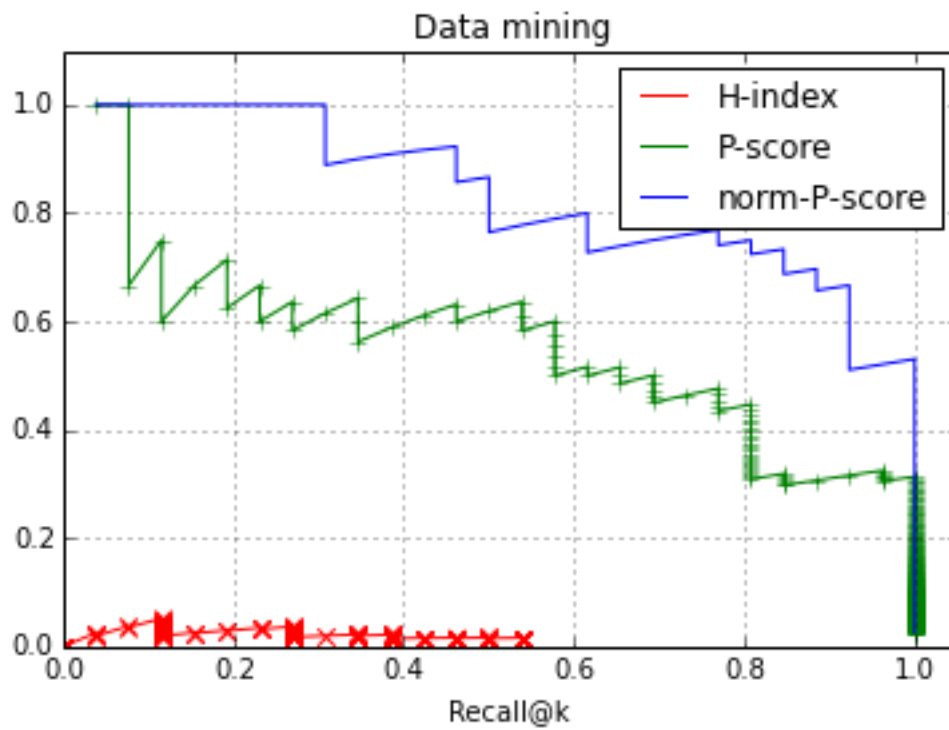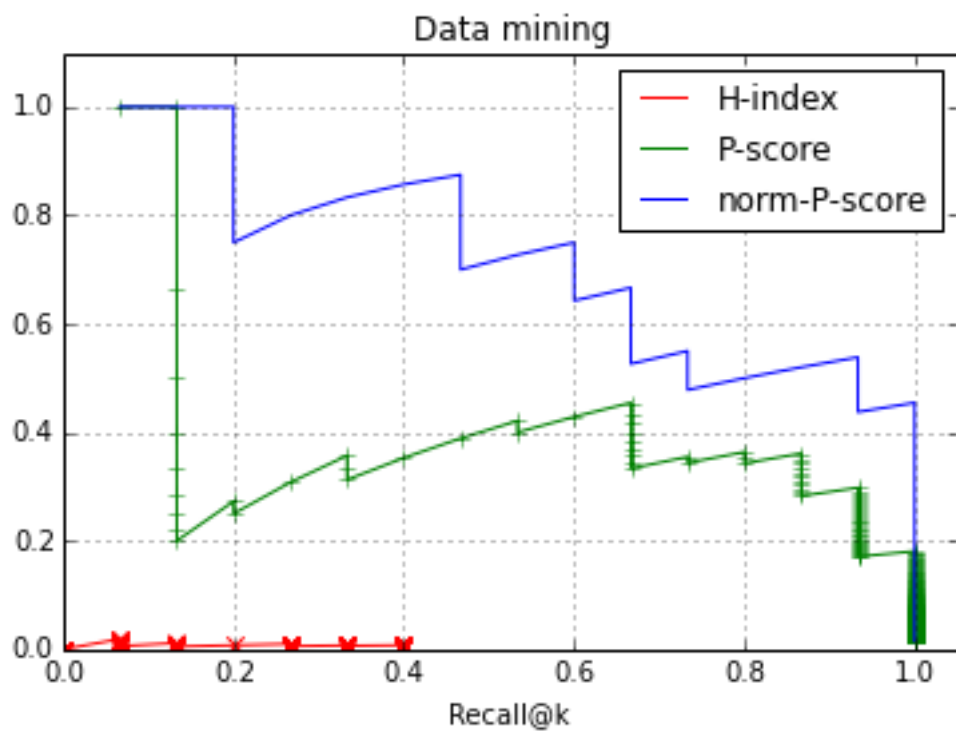
Information retrieval

In [133]: show_me('Database', 'db', 1)

Database
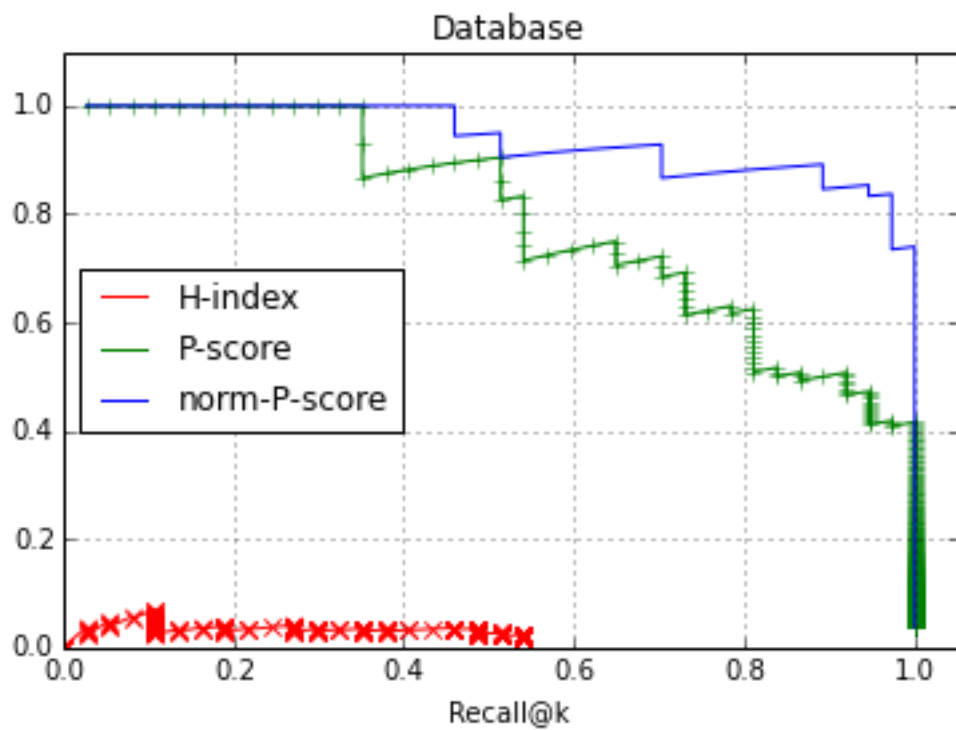
Data mining

Data mining

In [144]: show_me('Knowledge management', 'km', 2)

Knowledge management
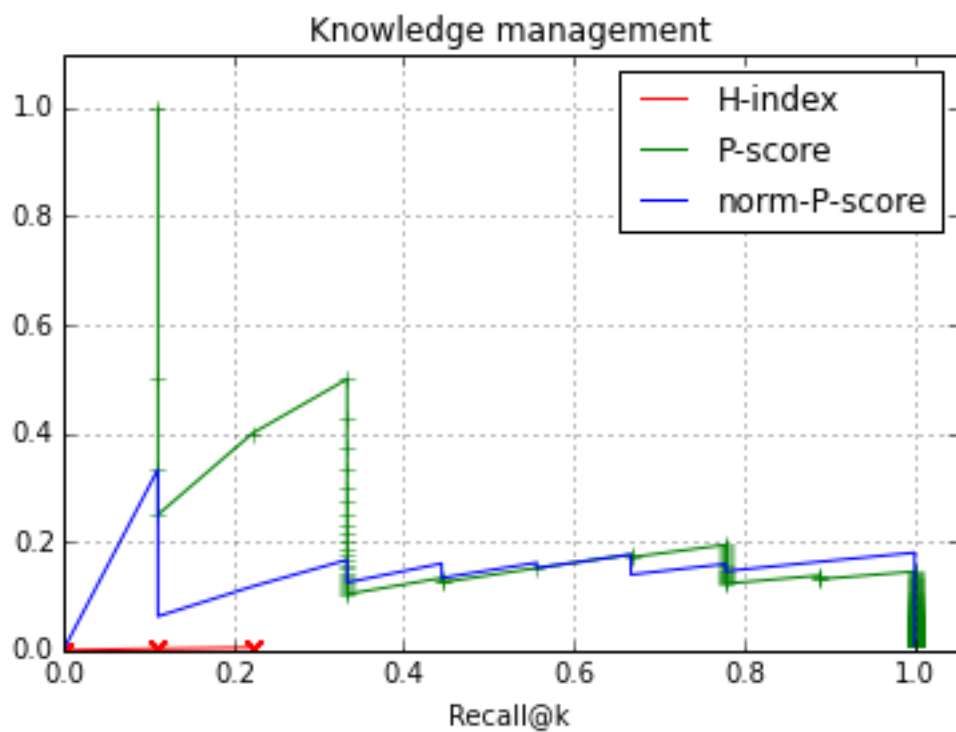
0.4 Plotting III

```
In [149]: pd.read_csv('expert-subareas-classification/ir-db-cn-berthier/Information retrieval_pr

Out[149]:       Unnamed: 0  Rank.0                Vkey.0  \
          0              0       1          journals/nar
          1              1       2             conf/balt
          2              2       3          journals/cce
          3              3       4             conf/cvpr
          4              4       5           journals/pr
          5              5       6          journals/cee
          6              6       7          journals/cma
          7              7       8   journals/neuroimage
          8              8       9             conf/mmar
          9              9      10            conf/icete
          10            10      11  journals/information
          11            11      12         journals/eswa
          12            12      13          journals/scl
          13            13      14   journals/automatica
          14            14      15           journals/cm
          15            15      16        journals/pieee
          16            16      17       journals/sigpro
          17            17      18              conf/www
          18            18      19         journals/www
          19            19      20          journals/or
          20            20      21       journals/ploscb
          21            21      22     conf/coordination
          22            22      23           conf/infocom
          23            23      24     journals/computer
          24            24      25        journals/bmcbi
          25            25      26           journals/ml
          26            26      27           journals/is
          27            27      28         journals/isci
          28            28      29            conf/dasip
          29            29      30         journals/bise
          ..           ...     ...                   ...
          970          970     971              conf/wia
          971          971     972           conf/wollic
          972          972     973          conf/secrypt
          973          973     974           conf/camad
          974          974     975              conf/fdl
          975          975     976           conf/icaris
          976          976     977             conf/adma
          977          977     978            conf/cacsd
          978          978     979             conf/wmte
```

979	979	980	conf/idaacs
980	980	981	conf/rsfdgrc
981	981	982	conf/ilp
982	982	983	conf/ems
983	983	984	conf/ic3
984	984	985	conf/icinco
985	985	986	conf/icsob
986	986	987	conf/medes
987	987	988	conf/mobilight
988	988	989	journals/at
989	989	990	conf/biosig
990	990	991	conf/mig
991	991	992	conf/epia
992	992	993	conf/mbmv
993	993	994	conf/csedu
994	994	995	conf/isvd
995	995	996	conf/wims
996	996	997	conf/eurocast
997	997	998	journals/cit
998	998	999	conf/asmta
999	999	1000	journals/jca

	Subareas.0	Precision.0@k	Recall.0@k \
0	NaN	0.000000	0.00
1	NaN	0.000000	0.00
2	NaN	0.000000	0.00
3	['Computer vision', 'Pattern recognition']	0.000000	0.00
4	NaN	0.000000	0.00
5	NaN	0.000000	0.00
6	NaN	0.000000	0.00
7	NaN	0.000000	0.00
8	NaN	0.000000	0.00
9	NaN	0.000000	0.00
10	NaN	0.000000	0.00
11	NaN	0.000000	0.00
12	NaN	0.000000	0.00
13	NaN	0.000000	0.00
14	NaN	0.000000	0.00
15	NaN	0.000000	0.00
16	NaN	0.000000	0.00
17	['World Wide Web']	0.000000	0.00
18	['World Wide Web']	0.000000	0.00
19	NaN	0.000000	0.00
20	NaN	0.000000	0.00
21	NaN	0.000000	0.00
22	['Computer network']	0.000000	0.00
23	['']	0.000000	0.00
24	NaN	0.000000	0.00

25	NaN	0.000000	0.00
26	['Database']	0.000000	0.00
27	['Knowledge management']	0.000000	0.00
28	NaN	0.000000	0.00
29	NaN	0.000000	0.00
..	...	...	...
970	NaN	0.011329	0.55
971	NaN	0.011317	0.55
972	NaN	0.011305	0.55
973	NaN	0.011294	0.55
974	NaN	0.011282	0.55
975	NaN	0.011270	0.55
976	NaN	0.011259	0.55
977	NaN	0.011247	0.55
978	NaN	0.011236	0.55
979	NaN	0.011224	0.55
980	NaN	0.011213	0.55
981	NaN	0.011202	0.55
982	NaN	0.011190	0.55
983	NaN	0.011179	0.55
984	NaN	0.011168	0.55
985	NaN	0.011156	0.55
986	NaN	0.011145	0.55
987	NaN	0.011134	0.55
988	NaN	0.011122	0.55
989	NaN	0.011111	0.55
990	NaN	0.011100	0.55
991	NaN	0.011089	0.55
992	NaN	0.011078	0.55
993	NaN	0.011066	0.55
994	NaN	0.011055	0.55
995	NaN	0.011044	0.55
996	NaN	0.011033	0.55
997	NaN	0.011022	0.55
998	NaN	0.011011	0.55
999	NaN	0.011000	0.55

	Rank.1	Vkey.1 \
0	1	conf/sigir
1	2	conf/cikm
2	3	conf/trec
3	4	conf/ecir
4	5	conf/clef
5	6	conf/www
6	7	journals/jasis
7	8	journals/ipm
8	9	journals/sigir
9	10	conf/mm

10	11	conf/aaai
11	12	conf/jcdl
12	13	journals/tois
13	14	journals/ir
14	15	conf/wsdm
15	16	conf/ntcir
16	17	conf/kdd
17	18	journals/tkde
18	19	conf/acl
19	20	conf/icdm
20	21	conf/spire
21	22	conf/icde
22	23	conf/ercimdl
23	24	conf/icmcs
24	25	conf/sigmod
25	26	conf/airs
26	27	conf/coling
27	28	conf/ijcai
28	29	conf/riao
29	30	conf/inex
..	...	...
970	971	conf/emc
971	972	journals/aes
972	973	journals/iandc
973	974	conf/seal
974	975	journals/ijcgt
975	976	conf/vlsi
976	977	conf/itng
977	978	journals/cal
978	979	conf/iasam
979	980	conf/arc
980	981	conf/itc
981	982	conf/sysose
982	983	conf/sitis
983	984	journals/mam
984	985	conf/iswc
985	986	journals/biodatamining
986	987	journals/ijwmc
987	988	conf/cf
988	989	journals/cars
989	990	journals/jvis
990	991	conf/icgec
991	992	conf/arith
992	993	conf/dils
993	994	conf/concur
994	995	conf/debs
995	996	journals/esi
996	997	journals/ijcnds

```
997      998           conf/bildmed
998      999         journals/jicce
999     1000              conf/sbp


                                           Subareas.1  Precision.1@k  \
0                          ['Information retrieval']       1.000000
1      ['Database', 'Information retrieval', 'Knowled...   1.000000
2                          ['Information retrieval']       1.000000
3                          ['Information retrieval']       1.000000
4                                             ['']         0.800000
5                               ['World Wide Web']         0.666667
6                          ['Information retrieval']       0.714286
7                          ['Information retrieval']       0.750000
8                          ['Information retrieval']       0.777778
9                                  ['Multimedia']         0.700000
10                      ['Artificial intelligence']       0.636364
11                         ['Information retrieval']       0.666667
12                         ['Information retrieval']       0.692308
13                         ['Information retrieval']       0.714286
14           ['Data mining', 'Information retrieval']      0.733333
15                         ['Information retrieval']       0.750000
16          ['Data mining', 'Knowledge management']       0.705882
17             ['Database', 'Knowledge management']       0.666667
18     ['Machine learning', 'Natural language process...   0.631579
19                               ['Data mining']         0.600000
20                         ['Information retrieval']       0.619048
21                                 ['Database']           0.590909
22                                             ['']         0.565217
23                                 ['Multimedia']         0.541667
24                                 ['Database']           0.520000
25                         ['Information retrieval']       0.538462
26              ['Natural language processing']         0.518519
27                      ['Artificial intelligence']       0.500000
28                         ['Information retrieval']       0.517241
29                         ['Information retrieval']       0.533333
..                                             ...             ...
970                                            NaN         0.020597
971                                            NaN         0.020576
972                                            NaN         0.020555
973                                            NaN         0.020534
974                                            NaN         0.020513
975                                            NaN         0.020492
976                                            NaN         0.020471
977                                            NaN         0.020450
978                                            NaN         0.020429
979                                            NaN         0.020408
980                                            NaN         0.020387
981                                            NaN         0.020367
```

		NaN	0.020346
982		NaN	0.020346
983		NaN	0.020325
984		NaN	0.020305
985		NaN	0.020284
986		NaN	0.020263
987		NaN	0.020243
988		NaN	0.020222
989		NaN	0.020202
990		NaN	0.020182
991		NaN	0.020161
992		NaN	0.020141
993		NaN	0.020121
994		NaN	0.020101
995		NaN	0.020080
996		NaN	0.020060
997		NaN	0.020040
998		NaN	0.020020
999		NaN	0.020000

	Recall.1@k	Rank.2	Vkey.2 \
0	0.05	1	conf/ictir
1	0.10	2	conf/sigir
2	0.15	3	conf/adcs
3	0.20	4	journals/ir
4	0.20	5	conf/ecir
5	0.20	6	conf/trec
6	0.25	7	journals/sigir
7	0.30	8	conf/iiix
8	0.35	9	journals/tois
9	0.35	10	conf/wsdm
10	0.35	11	conf/inex
11	0.40	12	conf/spire
12	0.45	13	conf/airs
13	0.50	14	conf/cikm
14	0.55	15	journals/tweb
15	0.60	16	conf/riao
16	0.60	17	conf/clef
17	0.60	18	conf/ntcir
18	0.60	19	conf/la-web
19	0.60	20	conf/jcdl
20	0.65	21	journals/tist
21	0.65	22	journals/ipm
22	0.65	23	conf/ercimdl
23	0.65	24	conf/www
24	0.65	25	conf/mir
25	0.70	26	conf/isi
26	0.70	27	conf/icwsm
27	0.70	28	conf/mmm

```
28        0.75      29            conf/adc
29        0.80      30           conf/icadl
..        ...       ...              ...
970       1.00      971         conf/icecsys
971       1.00      972       conf/edutainment
972       1.00      973          conf/iwinac
973       1.00      974           conf/mue
974       1.00      975           conf/ecms
975       1.00      976           conf/pdc
976       1.00      977          conf/walcom
977       1.00      978           conf/wsks
978       1.00      979           conf/hpsr
979       1.00      980          conf/sisap
980       1.00      981          conf/profes
981       1.00      982         conf/greencom
982       1.00      983           conf/vlsi
983       1.00      984           conf/mldm
984       1.00      985           conf/aime
985       1.00      986       conf/iconference
986       1.00      987          conf/cyberc
987       1.00      988          conf/ideal
988       1.00      989          conf/ciarp
989       1.00      990          conf/jurix
990       1.00      991          conf/synasc
991       1.00      992          conf/icgec
992       1.00      993           conf/iti
993       1.00      994          conf/quatic
994       1.00      995           conf/3dtv
995       1.00      996           conf/snpd
996       1.00      997           conf/dgci
997       1.00      998         journals/aml
998       1.00      999           conf/wwic
999       1.00      1000          conf/ispdc


                                        Subareas.2  Precision.2@k  \
0                      ['Information retrieval']       1.000000
1                      ['Information retrieval']       1.000000
2                                            NaN       0.666667
3                      ['Information retrieval']       0.750000
4                      ['Information retrieval']       0.800000
5                      ['Information retrieval']       0.833333
6                      ['Information retrieval']       0.857143
7     ['Human-computer interaction', 'Information re...       0.875000
8                      ['Information retrieval']       0.888889
9           ['Data mining', 'Information retrieval']       0.900000
10                     ['Information retrieval']       0.909091
11                     ['Information retrieval']       0.916667
12                     ['Information retrieval']       0.923077
```

13	['Database', 'Information retrieval', 'Knowled...	0.928571
14	NaN	0.866667
15	['Information retrieval']	0.875000
16	['']	0.823529
17	['Information retrieval']	0.833333
18	NaN	0.789474
19	['Information retrieval']	0.800000
20	NaN	0.761905
21	['Information retrieval']	0.772727
22	['']	0.739130
23	['World Wide Web']	0.708333
24	['Information retrieval', 'Multimedia']	0.720000
25	['Internet privacy']	0.692308
26	['World Wide Web']	0.666667
27	['Multimedia']	0.642857
28	NaN	0.620690
29	NaN	0.600000
..	...	...
970	NaN	0.020597
971	NaN	0.020576
972	NaN	0.020555
973	NaN	0.020534
974	NaN	0.020513
975	NaN	0.020492
976	NaN	0.020471
977	NaN	0.020450
978	NaN	0.020429
979	NaN	0.020408
980	NaN	0.020387
981	NaN	0.020367
982	NaN	0.020346
983	NaN	0.020325
984	NaN	0.020305
985	NaN	0.020284
986	NaN	0.020263
987	NaN	0.020243
988	NaN	0.020222
989	NaN	0.020202
990	NaN	0.020182
991	NaN	0.020161
992	NaN	0.020141
993	NaN	0.020121
994	NaN	0.020101
995	NaN	0.020080
996	NaN	0.020060
997	NaN	0.020040
998	NaN	0.020020
999	NaN	0.020000

	Recall.2@k
0	0.05
1	0.10
2	0.10
3	0.15
4	0.20
5	0.25
6	0.30
7	0.35
8	0.40
9	0.45
10	0.50
11	0.55
12	0.60
13	0.65
14	0.65
15	0.70
16	0.70
17	0.75
18	0.75
19	0.80
20	0.80
21	0.85
22	0.85
23	0.85
24	0.90
25	0.90
26	0.90
27	0.90
28	0.90
29	0.90
..	...
970	1.00
971	1.00
972	1.00
973	1.00
974	1.00
975	1.00
976	1.00
977	1.00
978	1.00
979	1.00
980	1.00
981	1.00
982	1.00
983	1.00
984	1.00

```
985      1.00
986      1.00
987      1.00
988      1.00
989      1.00
990      1.00
991      1.00
992      1.00
993      1.00
994      1.00
995      1.00
996      1.00
997      1.00
998      1.00
999      1.00

[1000 rows x 16 columns]
```
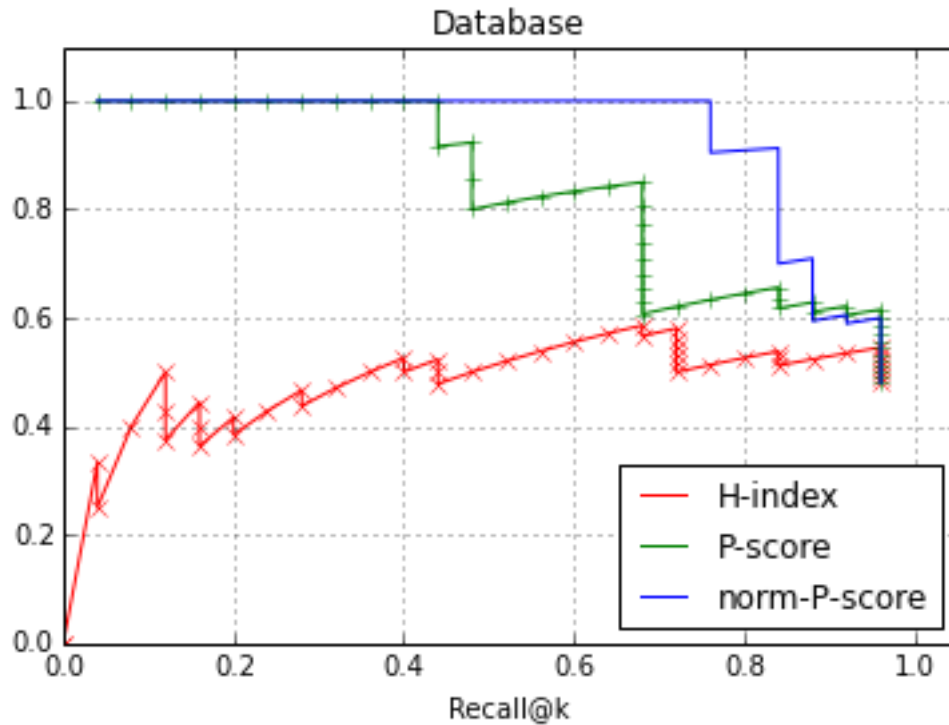
In [162]: show_me('Information retrieval', 'ir-db-cn-berthier', 2)
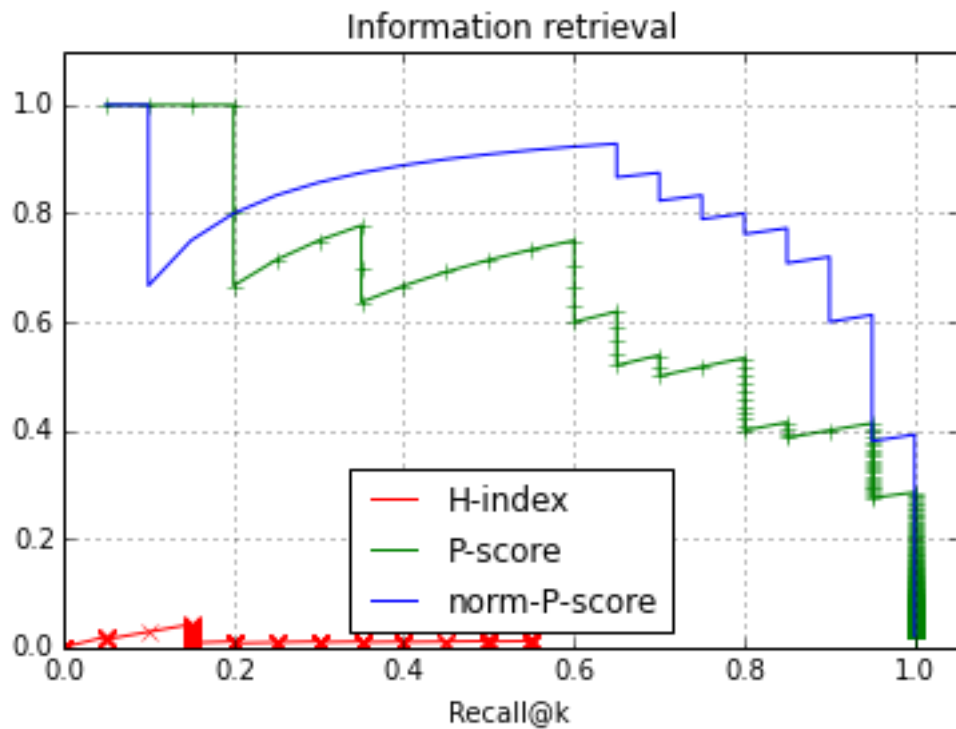
In [171]: show_me('Database', 'ir-db-cn-berthier', 2)

Database

In [169]: # TODO A few differences between the ones sent and the ones we get here
pd.read_csv('expert-subareas-classification/ir-db-cn-berthier/Information retrieval_pr

Out[169]:   Unnamed: 0 Rank.0   Vkey.0 \
0   0   1   conf/cvpr
1   1   2   conf/www
2   2   3   conf/icml
3   3   4   conf/nips
4   4   5   conf/sigmod
5   5   6   conf/wsdm
6   6   7   conf/emnlp
7   7   8   conf/icde
8   8   9   conf/sigir
9   9   10   journals/ir
10   10   11   conf/icwsm
11   11   12   conf/naacl
12   12   13   conf/ijcai
13   13   14   conf/aaai
14   14   15   conf/sdm
15   15   16   conf/vldb
16   16   17   conf/coling
17   17   18   conf/lrec
18   18   19   conf/clef

18

```
19      19      20         conf/icmcs
20      20      21         conf/webi
21      21      22         conf/jcdl
22      22      23         conf/spire
23      23      24         conf/trec
24      24      25          conf/mmm
25      25      26         conf/inex
26      26      27       journals/ipm
27      27      28     journals/sigir
28      28      29         conf/airs
29      29      30          conf/chi
30      30      31         conf/cikm
31      31      32         conf/ecir
32      32      33     journals/jasis
33      33      34           conf/mm
34      34      35      journals/tois
35      35      36        conf/ntcir
36      36      37          conf/kdd
37      37      38      journals/tkde
38      38      39          conf/acl
39      39      40         conf/icdm
40      40      41       conf/ercimdl
41      41      42         conf/riao
42      42      43          conf/isi
43      43      44      journals/cacm
44      44      45     journals/pvldb
45      45      46         conf/iiix
46      46      47        conf/ictir
47      47      48          conf/mir
48      48      49      journals/tmm
49      49      50          conf/dgo


                                     Subareas.0  Precision.0@k  \
0        ['Computer vision', 'Pattern recognition']       0.000000
1                                 ['World Wide Web']       0.000000
2                               ['Machine learning']       0.000000
3                               ['Machine learning']       0.000000
4                                      ['Database']       0.000000
5         ['Data mining', 'Information retrieval']       0.166667
6               ['Natural language processing']       0.142857
7                                      ['Database']       0.125000
8                        ['Information retrieval']       0.222222
9                        ['Information retrieval']       0.300000
10                                ['World Wide Web']       0.272727
11              ['Natural language processing']       0.250000
12                    ['Artificial intelligence']       0.230769
13                    ['Artificial intelligence']       0.214286
14                                   ['Data mining']       0.200000
```

19

15	['Database']	0.187500
16	['Natural language processing']	0.176471
17	['Natural language processing']	0.166667
18	['']	0.157895
19	['Multimedia']	0.150000
20	['World Wide Web']	0.142857
21	['Information retrieval']	0.181818
22	['Information retrieval']	0.217391
23	['Information retrieval']	0.250000
24	['Multimedia']	0.240000
25	['Information retrieval']	0.269231
26	['Information retrieval']	0.296296
27	['Information retrieval']	0.321429
28	['Information retrieval']	0.344828
29	['Human-computer interaction']	0.333333
30	['Database', 'Information retrieval', 'Knowled...	0.354839
31	['Information retrieval']	0.375000
32	['Information retrieval']	0.393939
33	['Multimedia']	0.382353
34	['Information retrieval']	0.400000
35	['Information retrieval']	0.416667
36	['Data mining', 'Knowledge management']	0.405405
37	['Database', 'Knowledge management']	0.394737
38	['Machine learning', 'Natural language process...	0.384615
39	['Data mining']	0.375000
40	['']	0.365854
41	['Information retrieval']	0.380952
42	['Internet privacy']	0.372093
43	['Algorithm']	0.363636
44	['Database']	0.355556
45	['Human-computer interaction', 'Information re...	0.369565
46	['Information retrieval']	0.382979
47	['Information retrieval', 'Multimedia']	0.395833
48	['Multimedia']	0.387755
49	['Algorithm']	0.380000

	Recall.0@k	Rank.1	Vkey.1 \
0	0.00	1	conf/sigir
1	0.00	2	conf/cikm
2	0.00	3	conf/trec
3	0.00	4	conf/ecir
4	0.00	5	conf/clef
5	0.05	6	conf/www
6	0.05	7	journals/jasis
7	0.05	8	journals/ipm
8	0.10	9	journals/sigir
9	0.15	10	conf/mm
10	0.15	11	conf/aaai

```
11      0.15    12      conf/jcdl
12      0.15    13      journals/tois
13      0.15    14      journals/ir
14      0.15    15      conf/wsdm
15      0.15    16      conf/ntcir
16      0.15    17      conf/kdd
17      0.15    18      journals/tkde
18      0.15    19      conf/acl
19      0.15    20      conf/icdm
20      0.15    21      conf/spire
21      0.20    22      conf/icde
22      0.25    23      conf/ercimdl
23      0.30    24      conf/icmcs
24      0.30    25      conf/sigmod
25      0.35    26      conf/airs
26      0.40    27      conf/coling
27      0.45    28      conf/ijcai
28      0.50    29      conf/riao
29      0.50    30      conf/inex
30      0.55    31      conf/icml
31      0.60    32      conf/nips
32      0.65    33      conf/vldb
33      0.65    34      conf/mmm
34      0.70    35      conf/isi
35      0.75    36      conf/webi
36      0.75    37      conf/naacl
37      0.75    38      journals/cacm
38      0.75    39      conf/chi
39      0.75    40      journals/pvldb
40      0.75    41      conf/iiix
41      0.80    42      conf/cvpr
42      0.80    43      conf/emnlp
43      0.80    44      conf/sdm
44      0.80    45      conf/ictir
45      0.85    46      conf/mir
46      0.90    47      journals/tmm
47      0.95    48      conf/icwsm
48      0.95    49      conf/dgo
49      0.95    50      conf/lrec


                                Subareas.1  Precision.1@k  \
0                       ['Information retrieval']      1.000000
1   ['Database', 'Information retrieval', 'Knowled...      1.000000
2                       ['Information retrieval']      1.000000
3                       ['Information retrieval']      1.000000
4                                           ['']      0.800000
5                           ['World Wide Web']      0.666667
6                       ['Information retrieval']      0.714286
```

7	['Information retrieval']	0.750000
8	['Information retrieval']	0.777778
9	['Multimedia']	0.700000
10	['Artificial intelligence']	0.636364
11	['Information retrieval']	0.666667
12	['Information retrieval']	0.692308
13	['Information retrieval']	0.714286
14	['Data mining', 'Information retrieval']	0.733333
15	['Information retrieval']	0.750000
16	['Data mining', 'Knowledge management']	0.705882
17	['Database', 'Knowledge management']	0.666667
18	['Machine learning', 'Natural language process...	0.631579
19	['Data mining']	0.600000
20	['Information retrieval']	0.619048
21	['Database']	0.590909
22	['']	0.565217
23	['Multimedia']	0.541667
24	['Database']	0.520000
25	['Information retrieval']	0.538462
26	['Natural language processing']	0.518519
27	['Artificial intelligence']	0.500000
28	['Information retrieval']	0.517241
29	['Information retrieval']	0.533333
30	['Machine learning']	0.516129
31	['Machine learning']	0.500000
32	['Database']	0.484848
33	['Multimedia']	0.470588
34	['Internet privacy']	0.457143
35	['World Wide Web']	0.444444
36	['Natural language processing']	0.432432
37	['Algorithm']	0.421053
38	['Human-computer interaction']	0.410256
39	['Database']	0.400000
40	['Human-computer interaction', 'Information re...	0.414634
41	['Computer vision', 'Pattern recognition']	0.404762
42	['Natural language processing']	0.395349
43	['Data mining']	0.386364
44	['Information retrieval']	0.400000
45	['Information retrieval', 'Multimedia']	0.413043
46	['Multimedia']	0.404255
47	['World Wide Web']	0.395833
48	['Algorithm']	0.387755
49	['Natural language processing']	0.380000

	Recall.1@k	Rank.2	Vkey.2 \
0	0.05	1	conf/ictir
1	0.10	2	conf/sigir
2	0.15	3	journals/ir

3	0.20	4	conf/ecir
4	0.20	5	conf/trec
5	0.20	6	journals/sigir
6	0.25	7	conf/iiix
7	0.30	8	journals/tois
8	0.35	9	conf/wsdm
9	0.35	10	conf/inex
10	0.35	11	conf/spire
11	0.40	12	conf/airs
12	0.45	13	conf/cikm
13	0.50	14	conf/riao
14	0.55	15	conf/clef
15	0.60	16	conf/ntcir
16	0.60	17	conf/jcdl
17	0.60	18	journals/ipm
18	0.60	19	conf/ercimdl
19	0.60	20	conf/www
20	0.65	21	conf/mir
21	0.65	22	conf/isi
22	0.65	23	conf/icwsm
23	0.65	24	conf/mmm
24	0.65	25	journals/jasis
25	0.70	26	conf/sdm
26	0.70	27	conf/mm
27	0.70	28	conf/kdd
28	0.75	29	journals/tkde
29	0.80	30	conf/naacl
30	0.80	31	journals/pvldb
31	0.80	32	conf/dgo
32	0.80	33	conf/emnlp
33	0.80	34	conf/icdm
34	0.80	35	conf/webi
35	0.80	36	conf/acl
36	0.80	37	journals/tmm
37	0.80	38	conf/icde
38	0.80	39	conf/vldb
39	0.80	40	conf/icml
40	0.85	41	conf/sigmod
41	0.85	42	conf/coling
42	0.85	43	conf/aaai
43	0.85	44	conf/ijcai
44	0.90	45	conf/icmcs
45	0.95	46	conf/nips
46	0.95	47	conf/lrec
47	0.95	48	conf/cvpr
48	0.95	49	journals/cacm
49	0.95	50	conf/chi

```
                                     Subareas.2  Precision.2@k  \
0                     ['Information retrieval']       1.000000
1                     ['Information retrieval']       1.000000
2                     ['Information retrieval']       1.000000
3                     ['Information retrieval']       1.000000
4                     ['Information retrieval']       1.000000
5                     ['Information retrieval']       1.000000
6   ['Human-computer interaction', 'Information re...  1.000000
7                     ['Information retrieval']       1.000000
8         ['Data mining', 'Information retrieval']     1.000000
9                     ['Information retrieval']       1.000000
10                    ['Information retrieval']       1.000000
11                    ['Information retrieval']       1.000000
12  ['Database', 'Information retrieval', 'Knowled...  1.000000
13                    ['Information retrieval']       1.000000
14                                         ['']        0.933333
15                    ['Information retrieval']       0.937500
16                    ['Information retrieval']       0.941176
17                    ['Information retrieval']       0.944444
18                                         ['']        0.894737
19                          ['World Wide Web']         0.850000
20        ['Information retrieval', 'Multimedia']      0.857143
21                         ['Internet privacy']       0.818182
22                          ['World Wide Web']         0.782609
23                             ['Multimedia']          0.750000
24                    ['Information retrieval']       0.760000
25                            ['Data mining']         0.730769
26                             ['Multimedia']          0.703704
27        ['Data mining', 'Knowledge management']     0.678571
28          ['Database', 'Knowledge management']      0.655172
29              ['Natural language processing']       0.633333
30                               ['Database']         0.612903
31                               ['Algorithm']        0.593750
32              ['Natural language processing']       0.575758
33                             ['Data mining']         0.558824
34                          ['World Wide Web']         0.542857
35  ['Machine learning', 'Natural language process...  0.527778
36                             ['Multimedia']          0.513514
37                               ['Database']         0.500000
38                               ['Database']         0.487179
39                         ['Machine learning']       0.475000
40                               ['Database']         0.463415
41              ['Natural language processing']       0.452381
42                    ['Artificial intelligence']     0.441860
43                    ['Artificial intelligence']     0.431818
44                             ['Multimedia']          0.422222
45                         ['Machine learning']       0.413043
46              ['Natural language processing']       0.404255
```

```
47              ['Computer vision', 'Pattern recognition']        0.395833
48                                          ['Algorithm']         0.387755
49                            ['Human-computer interaction']      0.380000


     Recall.2@k
0        0.05
1        0.10
2        0.15
3        0.20
4        0.25
5        0.30
6        0.35
7        0.40
8        0.45
9        0.50
10       0.55
11       0.60
12       0.65
13       0.70
14       0.70
15       0.75
16       0.80
17       0.85
18       0.85
19       0.85
20       0.90
21       0.90
22       0.90
23       0.90
24       0.95
25       0.95
26       0.95
27       0.95
28       0.95
29       0.95
30       0.95
31       0.95
32       0.95
33       0.95
34       0.95
35       0.95
36       0.95
37       0.95
38       0.95
39       0.95
40       0.95
41       0.95
42       0.95
```
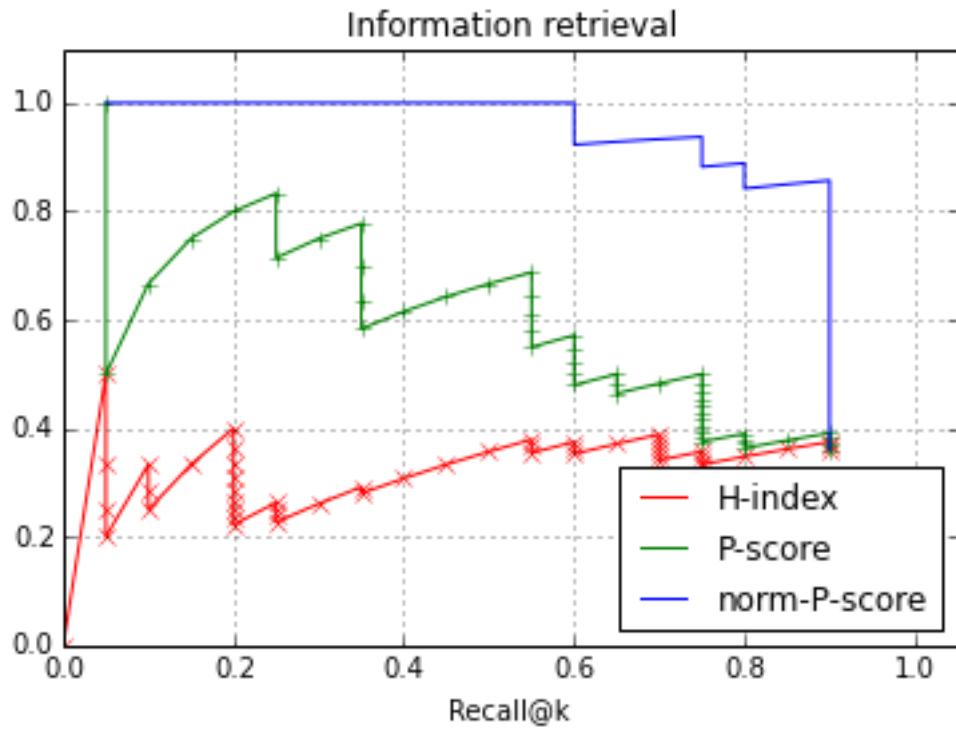
```
43          0.95
44          0.95
45          0.95
46          0.95
47          0.95
48          0.95
49          0.95
```

In [173]: show_me('Information retrieval', 'ir-students', 2)

Information retrieval

In [174]: show_me('Database', 'db', 1)

Database
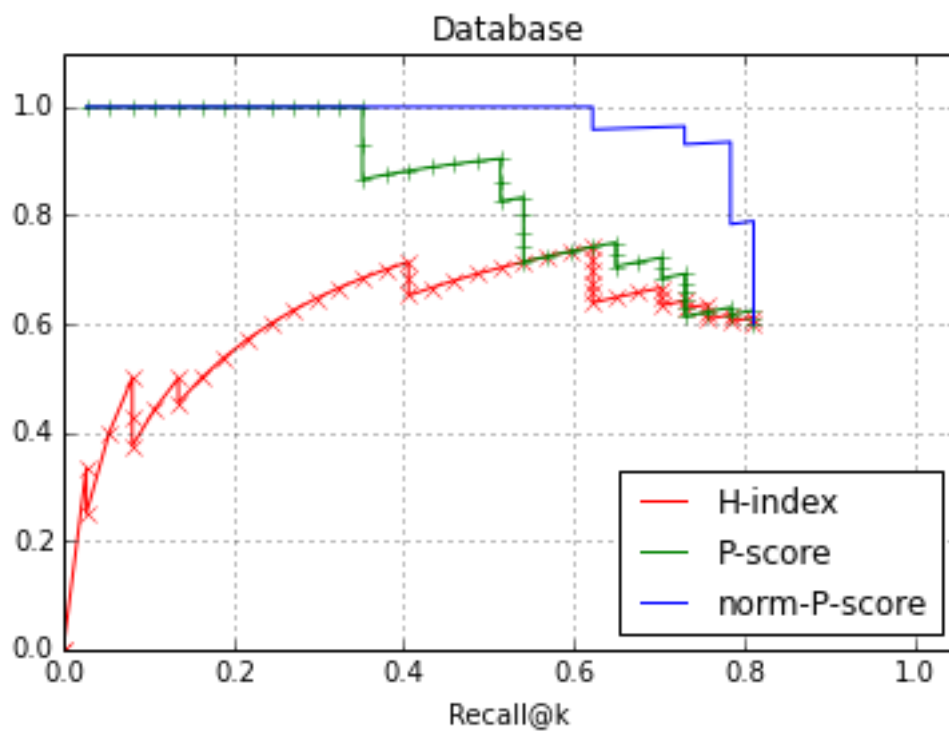
In [177]: show_me('Knowledge management', 'km', 2)

Knowledge management
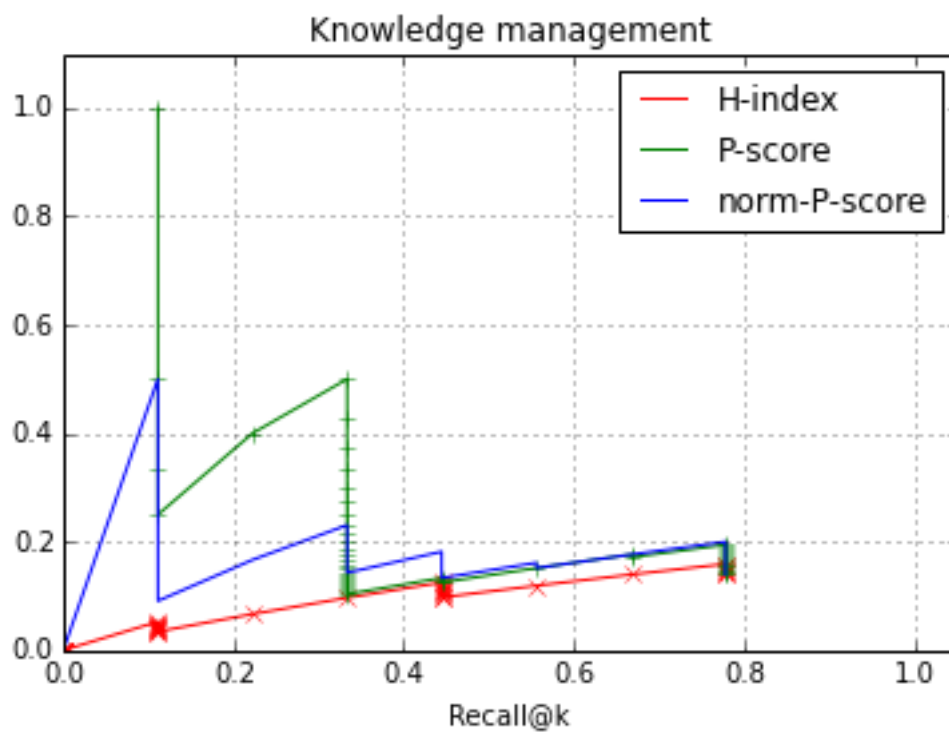
In [180]: show_me('Knowledge management', 'km', 1)

Knowledge management
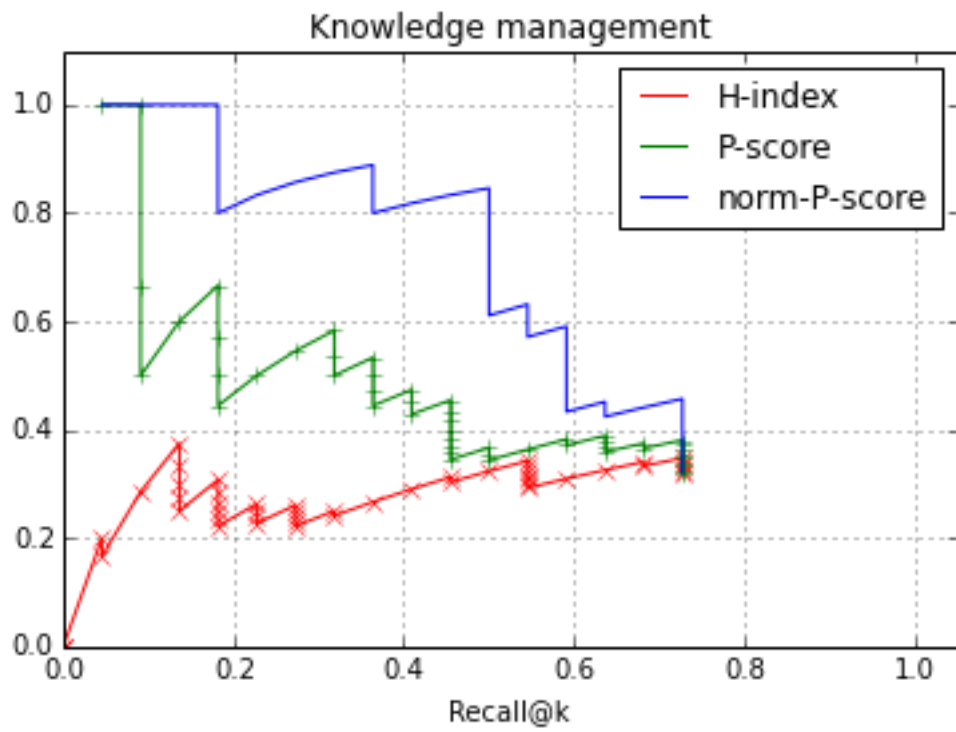
In [179]: show_me('Data mining', 'dm', 2)

Data mining
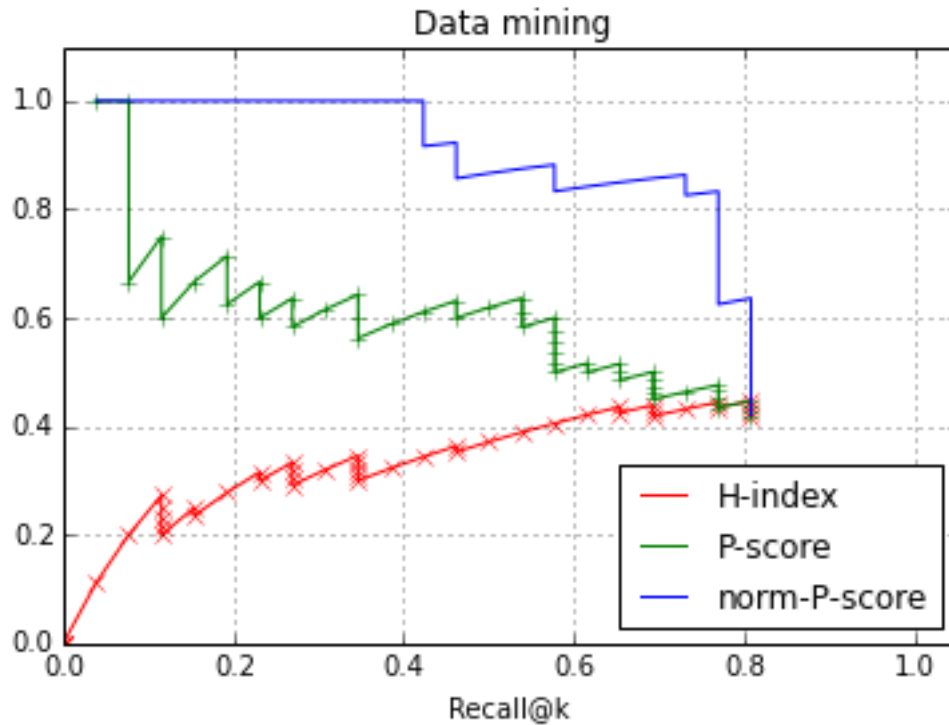
0.5 Listing the differences

```
In [235]: def list_diff(dirname, subarea, min_votes):
              sent = pd.read_csv('expert-subareas-classification/' + dirname + '/venues_subareas
              rank = pd.DataFrame(pd.read_csv('input/march-7-50-venues/' + subarea + '.csv')['Vk

              m = sent.merge(rank, left_on='Vkey', right_on='Vkey.2', how='outer')
              m = m.rename(columns={'Vkey': 'Sent', 'Vkey.2': 'Ranked'})
              m = m[m.isnull().any(axis=1)]
              m = m.reset_index(drop=True)

              m.to_csv('expert-subareas-classification/differences/diff-' + dirname + '.csv')

              return m

          ir = list_diff('ir-students', 'Information retrieval', 2)
          db = list_diff('db', 'Database', 1)
          dm = list_diff('dm', 'Data mining', 2)
          km = list_diff('km', 'Knowledge management', 2)

In [236]: dm

Out[236]:                  Sent                                      Subareas  \
          0    journals/tist                    ['Artificial intelligence']
```

1	journals/sadm	['Data mining', 'Data science']
2	conf/ssd	['Database']
3	conf/ilp	['Artificial intelligence']
4	conf/aistats	['Artificial intelligence', 'Data mining', 'Ma...
5	conf/icwsm	['Data science', 'World Wide Web']
6	conf/dawak	['Data mining', 'Database']
7	journals/tois	['Data science']
8	journals/ida	['Artificial intelligence', 'Data mining', 'Ma...
9	journals/jiis	['Artificial intelligence', 'Knowledge managem...
10	journals/mms	['Data science', 'Multimedia']
11	journals/tods	['Database']
12	conf/mdm	['Data science', 'Knowledge management']
13	conf/emnlp	['Natural language processing']
14	conf/gis	['Data science', 'Database']
15	conf/waim	['Data mining', 'Information retrieval', 'Know...
16	NaN	['']
17	NaN	NaN
18	NaN	NaN
19	NaN	NaN
20	NaN	NaN
21	NaN	NaN
22	NaN	NaN
23	NaN	NaN
24	NaN	NaN
25	NaN	NaN
26	NaN	NaN
27	NaN	NaN
28	NaN	NaN
29	NaN	NaN
30	NaN	NaN
31	NaN	NaN
32	NaN	NaN

|	Ranked
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
5	NaN
6	NaN
7	NaN
8	NaN
9	NaN
10	NaN
11	NaN
12	NaN
13	NaN

```
14                      NaN
15                      NaN
16                      NaN
17                conf/webi
18                conf/ijcai
19                  conf/mm
20               conf/icmcs
21             journals/tnn
22                 conf/acl
23                 conf/sac
24                conf/cvpr
25            journals/ijon
26            journals/isci
27           journals/bmcbi
28                conf/ipps
29  journals/bioinformatics
30              journals/pr
31              conf/igarss
32                conf/icip
```
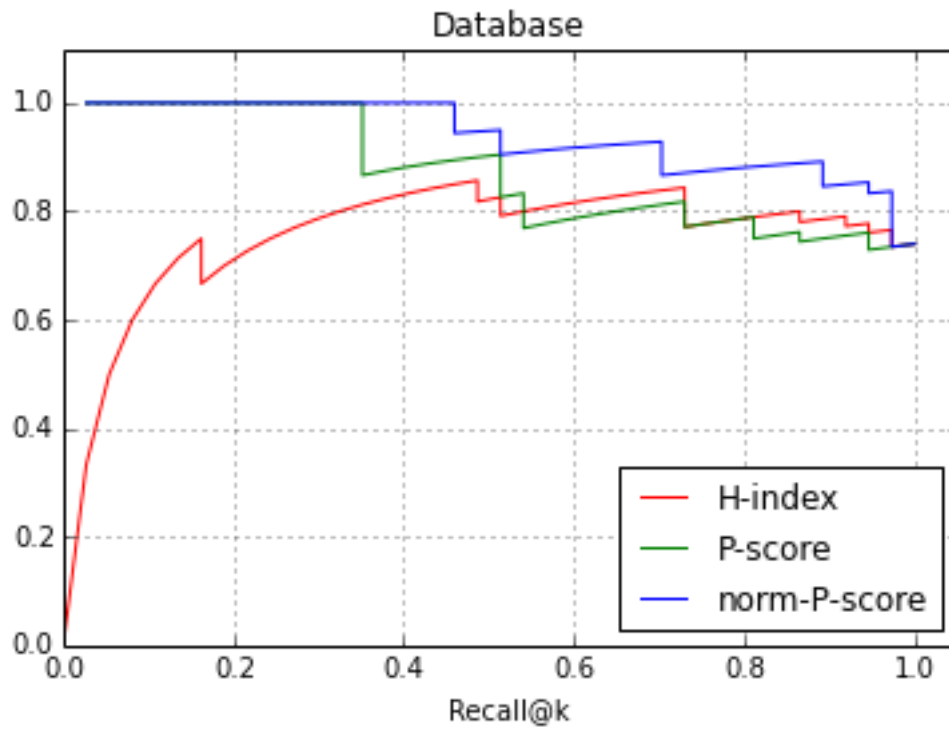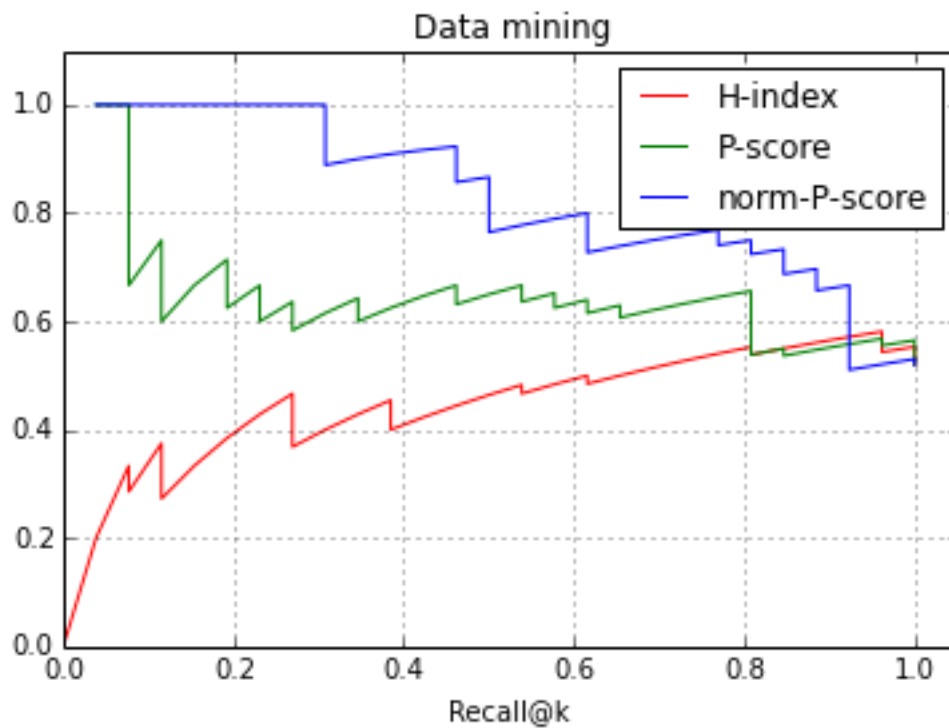
0.6 Plotting IV

```
In [266]: show_me('Information retrieval', 'ir-students', 2)
```


Information retrieval
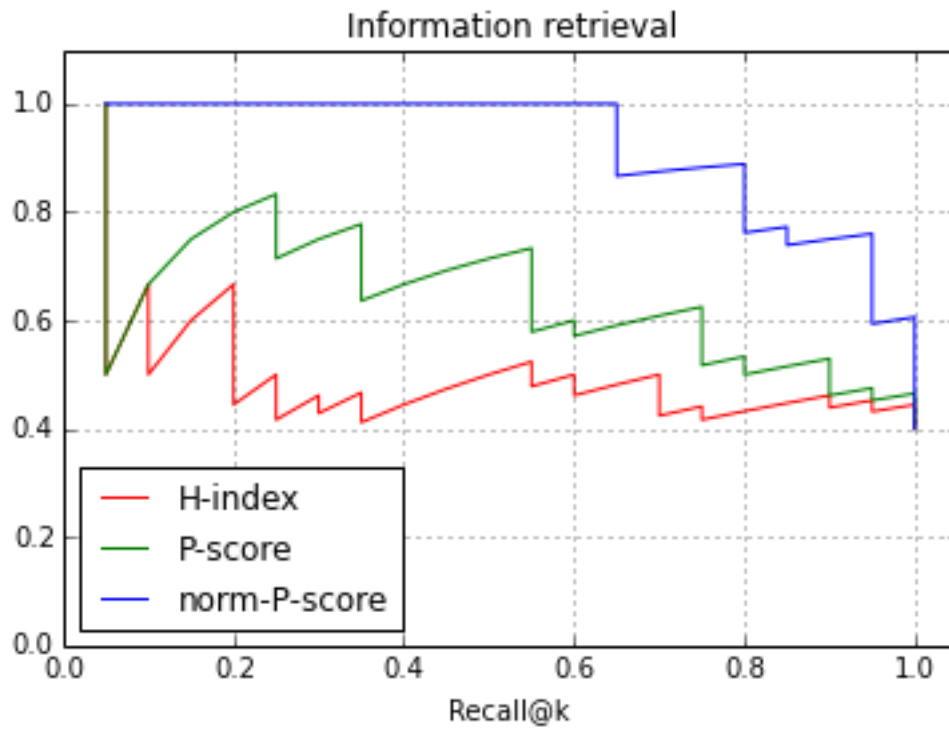
In [267]: show_me('Database', 'db', 1)

Database

Recall@k

In [268]: show_me('Data mining', 'dm', 2)

Data mining

Recall@k

In [269]: show_me('Knowledge management', 'km', 2)

Knowledge management