

## Precision / Recall:

```
*****
** Precision / Recall averaged over 33 queries **
*****
```

Permutation Name	P 0.25	P 0.50	P 0.75	P 1.00	P mean1	P mean2	P norm	R norm
=====	=====	=====	=====	=====	=====	=====	=====	=====
RAW TF	0.367	0.241	0.118	0.047	0.242	0.297	0.630	0.904
Boolean	0.179	0.114	0.061	0.016	0.118	0.131	0.532	0.883
Dice	0.481	0.346	0.222	0.099	0.350	0.390	0.713	0.931
Jaccard	0.262	0.137	0.032	0.011	0.144	0.185	0.448	0.616
Overlap	0.540	0.347	0.213	0.083	0.367	0.427	0.716	0.933
Unstem	0.514	0.270	0.114	0.053	0.300	0.370	0.614	0.827
No stwd	0.541	0.385	0.247	0.081	0.391	0.429	0.730	0.937
Region 1111	0.546	0.362	0.242	0.086	0.383	0.439	0.733	0.933
Region 1114	0.456	0.297	0.178	0.061	0.310	0.353	0.700	0.929
Default	0.549	0.385	0.247	0.098	0.394	0.436	0.732	0.932

## Precision / Recall with bigrams

```
*****
** Precision / Recall averaged over 33 queries **
*****
```

Permutation Name	P 0.25	P 0.50	P 0.75	P 1.00	P mean1	P mean2	P norm	R norm
=====	=====	=====	=====	=====	=====	=====	=====	=====
RAW TF	0.429	0.261	0.127	0.052	0.273	0.340	0.642	0.907
Boolean	0.220	0.146	0.079	0.018	0.148	0.183	0.554	0.888
Dice	0.525	0.351	0.242	0.085	0.372	0.431	0.725	0.932
Jaccard	0.343	0.168	0.107	0.041	0.206	0.288	0.553	0.733
Overlap	0.555	0.361	0.223	0.090	0.380	0.429	0.728	0.934
Unstem	0.470	0.284	0.114	0.055	0.290	0.360	0.616	0.828
No stwd	0.588	0.405	0.257	0.085	0.417	0.476	0.741	0.937
Region 1111	0.583	0.410	0.250	0.091	0.414	0.467	0.747	0.934
Region 1114	0.539	0.347	0.202	0.060	0.363	0.416	0.718	0.931
Default	0.565	0.385	0.255	0.095	0.402	0.461	0.743	0.933

## Interactive Query:

```
pan@pan: ~/test/hw2

Query (1): i am interested in security
Query (2):

Please enter the names of any authors you wish to search
for, one per line. Press [Enter] on a line by itself when
you're finished:

Author (1):

Saving query to 'interactive.raw'
Tokenizing and stemming query.
Making histogram of the query.

.I 1
.W
i
am
interest
in
secur

*****
Documents Most Similar To Interactive Query number 0
*****
Similarity Doc# Author Title
=====
0.51792873 2870 Denning # A Lattice Model of Secure
0.47458022 3068 Popek # A Model for Verification o
0.47031841 3111 Merkle # Secure Communications Over
0.43593659 2945 Denning # Certification of Programs
0.40756012 2372 Conway # On the Implementation of S
0.40234478 3174 Morris # Password Security: A Case
0.37749590 2621 Purdy # A High Security Log-in Pro
0.37301612 2869 Millen # Security Kernel Validation
0.36901610 3158 Denning # Secure Personal Computing
0.33962520 356 Ingeman # INTEREST (Algorithm 45)
0.33439138 740 Wright # INTEREST (Algorithm 45)
0.28099612 136 Ingeman # A Note on the Calculation
0.22272487 2622 Evans # A User Authentication Sche
0.18546255 1808 Van # Advanced Cryptographic Tec
0.17849473 2436 Lampson # A Note on the Confinement
0.17458949 2632 Wulf # HYDRA: The Kernel of a Mul
0.16318048 1746 Graham # Protection in an Informati
```

## Bigram Improvement:

As you can see in the screen shot, we used bigram sets in both queries:

In the left part we used the **stanford university** through the interactive query.

In the right part we used the **university stanford** through the interactive query.

The document #3204 has the bigrams **stanford university**, thus it ranks highly in the left part, whereas in the right part it ranks lower because we don't have the bigram **university stanford** in our document.

**Left Terminal Window (Query: stanford university):**

```
INTERACTIVE QUERY:
Please enter your query. Press [Enter] on a line by itself to finish entering your query:
Query (1): stanford university
Query (2):

Please enter the names of any authors you wish to search for, one per line. Press [Enter] on a line by itself when you're finished:
Author (1):

Saving query to 'interactive.raw'
Tokenizing and stemming query.
Making histogram of the query.

.I 1
.W
stanford
univers
stanford+univers

*****
Documents Most Similar To Interactive Query number 0
*****
Similarity Doc# Author Title
*****
0.31035675 1654 Forsythe # A University's Educational
0.20072455 1862 Finerman # Computing Capabilities at
0.16882808 3204 Korsvold # An On-Line Program for Non
0.12055956 68 Fein # The Role of the University
0.11696579 171 Reeves # Digital Computers in Unive
0.11483187 157 Reeves # Digital Computers in Unive
0.11385924 148 Reeves # Digital Computers in Unive
0.11294447 135 Reeves # Digital Computers in Unive
0.08870461 1349 Finerman # Computing Capabilities at
0.08620066 147 -- # Report on a Conference of
0.08512434 1100 Atchison # Status of Computer Science
0.08402429 382 Bush # Statistical Programs at th
0.07557074 188 Berezin # The Department of Computer
0.05879539 1533 Foley # A Markovian Model of the Un
0.05416872 1472 Lurch # Description of a High Capa
```

**Right Terminal Window (Query: university stanford):**

```
INTERACTIVE QUERY:
Please enter your query. Press [Enter] on a line by itself to finish entering your query:
Query (1): university stanford
Query (2):

Please enter the names of any authors you wish to search for, one per line. Press [Enter] on a line by itself when you're finished:
Author (1):

Saving query to 'interactive.raw'
Tokenizing and stemming query.
Making histogram of the query.

.I 1
.W
univers
stanford
univers+stanford

*****
Documents Most Similar To Interactive Query number 0
*****
Similarity Doc# Author Title
*****
0.32177626 1654 Forsythe # A University's Educational
0.26632913 1862 Finerman # Computing Capabilities at
0.15996310 68 Fein # The Role of the University
0.15519474 171 Reeves # Digital Computers in Unive
0.15236338 157 Reeves # Digital Computers in Unive
0.15001139 148 Reeves # Digital Computers in Unive
0.14905011 135 Reeves # Digital Computers in Unive
0.12663790 3204 Korsvold # An On-Line Program for Non
0.11769672 1349 Finerman # Computing Capabilities at
0.11437438 147 -- # Report on a Conference of
0.11294028 1100 Atchison # Status of Computer Science
0.11148669 382 Bush # Statistical Programs at th
0.10027018 188 Berezin # The Department of Computer
0.07801201 1533 Foley # A Markovian Model of the Un
```

**Padre Window (Document #3204):**

```
4 | vector1.pri | gen_bigrams.pri | query.stemmed | cacm.stem
42619 to the present version, it has proved quite useful indeed.
42620 .A
42621 Fletcher, J. G.
42622 .I 3204
42623 .T
42624 An On-Line Program for Non-Numerical Algebra
42625 .W
42626 The goal of this program is to make a step toward to design
42627 mathematical assistant. Some requirements for such a program
42628 easy to access, and that the result must be obtained in a rea
42629 time. Accordingly the program is written for a time-shared co
42630 computer as System Development Corporation, Santa Monica, Cal
42631 chosen because it also had a LISP 1.5 compiler. Programming is
42632 done from a remote teletype console at Stanford University.
42633 .A
42634 Korsvold, K.
42635 .I 0
42636
```