100 MOST FREQUENT WORDS

```
  [1] "edu"               "office"            "rainbowthreedigit"
  [4] "rainbowtwodigit"   "rainbowdigits"     "rainbowonedigit"
  [7] "rainbowtime"       "are"               "in"
 [10] "other"             "can"               "also"
 [13] "be"                "by"                "students"
 [16] "course"            "home"              "page"
 [19] "algorithms"        "this"              "the"
 [22] "for"               "and"               "we"
 [25] "will"              "an"                "of"
 [28] "data"              "as"                "design"
 [31] "with"              "from"              "some"
 [34] "information"       "links"             "more"
 [37] "homework"          "you"               "mail"
 [40] "cs"                "rainbowphonenum"   "is"
 [43] "that"              "to"                "all"
 [46] "one"               "not"               "class"
 [49] "lecture"           "may"               "at"
 [52] "have"              "on"                "software"
 [55] "program"           "last"              "rainbowfourdigit"
 [58] "university"        "computer"          "science"
 [61] "department"        "fall"              "programming"
 [64] "group"             "language"          "new"
 [67] "your"              "here"              "or"
 [70] "if"                "email"             "time"
 [73] "postscript"        "student"           "lab"
 [76] "project"           "am"                "using"
 [79] "it"                "professor"         "system"
 [82] "which"             "work"              "performance"
 [85] "phone"             "he"                "about"
 [88] "engineering"       "systems"           "based"
 [91] "computing"         "web"               "has"
 [94] "research"          "my"                "me"
 [97] "parallel"          "distributed"       "rainbowfivedigit"
[100] "conference"        "interests"
```

"NON INFORMATIVE" WORDS

```
 [1] "edu"   "are"   "in"    "other" "can"   "also"  "be"    "by"    "this"
[10] "the"   "for"   "and"   "we"    "will"  "an"    "of"    "as"    "with"
[19] "from"  "some"  "more"  "you"   "mail"  "cs"    "is"    "that"  "to"
[28] "all"   "not"   "may"   "at"    "have"  "on"    "your"  "or"    "if"
[37] "email" "am"    "it"    "which" "he"    "about" "web"   "has"
```

CONFUSION MATRICES BASED ON 57 WORDS

- RAINBOW *(Souped up Naive Bayes Classifier)*
  Rates nearly equivalent to those using 2,000 words

Classified as

		Students	Projects	Faculties	Courses	Totals
True	Students	144	4	10	5	163
	Projects	2	16	7	0	25
	Faculties	9	0	32	1	42
	Courses	8	8	3	54	73
	Totals	163	28	52	60	303

	Students	Projects	Faculties	Courses
Accuracy	88%	64%	76%	74%
Coverage	88%	57%	62%	90%

- Linear Discriminant Model *(four classes)*

Classified as

		Students	Projects	Faculties	Courses	Totals
True	Students	159	1	6	1	167
	Projects	10	15	0	1	26
	Faculties	10	2	34	0	46
	Courses	20	1	0	52	73
	Totals	199	19	40	54	312

	Students	Projects	Faculties	Courses
Accuracy	95%	58%	74%	71%
Coverage	80%	79%	85%	99%

- Mixed Nested Model *(Logistic [student vs other] and LDA [remaining three classes])*

Classified as

		Students	Projects	Faculties	Courses	Totals
True	Students	152	5	9	0	166
	Projects	6	9	8	3	26
	Faculties	6	2	32	6	46
	Courses	2	2	3	66	73
	Totals	166	18	52	75	311

	Students	Projects	Faculties	Courses
Accuracy	92%	33%	69%	90%
Coverage	91%	52%	6%1	88%