

SAS-L Counts

Howard Schreier
U.S. Dept. of Commerce

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. © indicates USA. Other brand and product names are registered trademarks or Trademarks of their respective companies.

Disclaimers

- Quantity is not quality.
- There are errors and omissions. I hope and believe they are minor.

Year ending 31 March 2004

Ron Fehd	726	Chang Chung	284
Howard Schreier	630	Harry Droogendyk	274
David Cassell	475	Dale McLerran	272
Richard DeVenezia	450	Peter Crawford	241
Jack Hamilton	421	Roland Rashleigh-Berry	215
Paul Choate	403	Paul Dorfman	214
Ian Whitlock	362	Venky Chakravarthy	205
Jim Groeneveld	343	Dennis Diskin	199
Sig Hermansen	305	Prasad Ravi	194
Ya Huang	288	Chuck Patridge	190

Cumulative, through 31 March 2004

Peter Crawford	2,715	Bill Viergever	1,618
Jack Hamilton	2,666	Richard DeVenezia	1,544
Ron Fehd	2,535	Jim Groeneveld	1,502
John Whittington	2,158	Ya Huang	1,460
Paul Dorfman	2,117	Gerhard Hellriegel	1,402
Ian Whitlock	2,107	Chuck Patridge	1,335
Howard Schreier	2,088	Sig Hermansen	1,325
David Cassell	1,935	Dale McLerran	1,203
Tim Berryhill	1,842	Bernard Tremblay	889
Karsten Self	1,778	Bob Hamer	861

Days (count only one per day)

Jack Hamilton	1,369	Sig Hermansen	818
Howard Schreier	1,193	Bill Viergever	771
Ian Whitlock	1,157	Chuck Patridge	703
Peter Crawford	1,110	Ya Huang	677
Paul Dorfman	1,076	David Cassell	665
John Whittington	1,049	Dale McLerran	637
Tim Berryhill	1,034	Bob Hamer	616
Karsten Self	997	Melvin Klassen	568
Richard DeVenezia	902	Mike Rhoads	559
Ron Fehd	848	Phil Miller	537

Continuity / persistence

- Unit: calendar quarter during which an individual posts five or more times
- Consecutive quarters
 - “Live” streaks
 - Other streaks
- Numerous nonconsecutive quarters

Streaks continuing through 2004-Q1

- Jack Hamilton: 42 quarters
- John Whittington: 37
- Joe Kelley: 35
- Sig Hermansen: 29
- Peter Crawford: 28
- Paul Dorfman: 28
- Bill Viergever: 27

7

Other long streaks

- Ian Whitlock: 43 quarters, through 2003-Q4
- Bob Hamer: 37, through 1997-Q2
- Melvin Klassen: 34, through 1998-Q3
- Karsten Self: 33, through 2002-Q1
- Phil Gallagher: 33, through 1996-Q3
- Phil Miller: 30, through 1994-Q2
- Tim Berryhill: 30, through 2000-Q1

8

Others with many qualifying quarters

- Howard Schreier: 48 quarters
- Mike Rhoads: 42
- Ray Pass: 39
- Michael Davis: 38
- Michael Friendly: 38
- Bob Abelson: 38
- Paul Thompson, Bob Schechter, Jules Bosch: 34

9

Year by year

- Top 7 by year, 1987-2003
- ↑ First appearance in top 7
- ↓ Last appearance in top 7

10

Year by year

1987

↑ Nelson Pardee	69
↑ Phil Miller	65
↑ Bob Parks	60 ↓
↑ June Genis	58 ↓
↑ Barry Grau	19
↑ Kevin Broekhoven	18 ↓
↑ Landy Manderson	16 ↓

11

Year by year

1988

Phil Miller	90
↑ Hal German	86
Nelson Pardee	67
↑ Bob Hamer	66
Barry Grau	62
↑ Pete Rikard	54 ↓
↑ Frank DiIorio	34 ↓

12

Year by year



1989

	Bob Hamer	165
	Phil Miller	163
↑	Howard Schreier	115
	Hal German	94 ↓
↑	Phil Gallagher	72
	Barry Grau	62
	Nelson Pardee	54 ↓

13

Year by year



1990

	Bob Hamer	198
	Phil Miller	117
	Howard Schreier	74
	Phil Gallagher	66
↑	Jack Hamilton	59
↑	Deb Cannon	55 ↓
	Barry Grau	55 ↓

14

Year by year



1991

	Phil Miller	93
↑	Don MacQueen	74 ↓
↑	Evan Cooch	73 ↓
	Jack Hamilton	72
↑	Sally Muller	70
	Phil Gallagher	65
	Bob Hamer	61 ↓

15

Year by year



1992

	Howard Schreier	148
↑	Kernon Gibes	103
	Phil Miller	83 ↓
↑	Melvin Klassen	78
↑	Bob Snyder	73
↑	Paul Oldenkamp	70 ↓
	Sally Muller	63 ↓

16

Year by year



1993

↑	Tim Berryhill	113
	Bob Snyder	102 ↓
↑	Ian Whitlock	96
↑	Tom Skinner	84
↑	Jay Wind	72 ↓
	Howard Schreier	72
	Kernon Gibes	69 ↓

17

Year by year



1994

	Tim Berryhill	238
	Ian Whitlock	150
↑	Kent Drew Spencer	145 ↓
	Phil Gallagher	124 ↓
	Jack Hamilton	124
	Tom Skinner	121 ↓
↑	Phil Mason	114

18

Year by year

1995

↑ John Whittington	277
Phil Mason	228 ↓
Tim Berryhill	220
Melvin Klassen	124 ↓
Ian Whitlock	124
Jack Hamilton	113
↑ Andrew Cary	104

19

Year by year

1996

Tim Berryhill	359
John Whittington	338
↑ Karsten Self	209
Andrew Cary	195 ↓
↑ Don Stanley	125 ↓
↑ Gerald Zuckier	121 ↓
Ian Whitlock	113

20

Year by year

1997

Tim Berryhill	330
↑ Anthony Ayiomamitis	235 ↓
↑ Bernard Tremblay	221
Karsten Self	168
John Whittington	142
↑ Jay Weedon	126 ↓
↑ Sig Hermansen	114 ↓

21

Year by year

1998

Karsten Self	282
↑ Peter Crawford	279
Tim Berryhill	258
Jack Hamilton	201
Bernard Tremblay	177
↑ Paul Dorfman	163
John Whittington	147

22

Year by year

1999

Peter Crawford	419
John Whittington	383 ↓
↑ Bill Viergever	311
Jack Hamilton	263
Bernard Tremblay	223 ↓
Tim Berryhill	201 ↓
Karsten Self	193

23

Year by year

2000

Paul Dorfman	535
↑ Gerhard Hellriegel	506 ↓
Peter Crawford	454
↑ Jim Groeneveld	346
↑ Ron Fehd	342
Jack Hamilton	335
Bill Viergever	323

24

Year by year

2001

Peter Crawford	436	
Karsten Self	422	↓
Bill Viergever	396	↓
↑ Chuck Patridge	383	↓
↑ Ya Huang	366	
Paul Dorfman	359	
↑ David Cassell	345	

25

Year by year

2002

Ron Fehd	745	
David Cassell	704	
Jim Groeneveld	615	↓
Peter Crawford	592	
Paul Dorfman	524	↓
Ian Whitlock	469	
Jack Hamilton	444	

26

Year by year

2003

Ron Fehd	695	↓
David Cassell	619	↓
Howard Schreier	486	↓
Ian Whitlock	480	↓
Peter Crawford	413	↓
Ya Huang	401	↓
Jack Hamilton	398	↓

27

SAS-L archives

- 1986 to present
- ~200 thousand postings
- >20 thousand distinct e-mail addresses
- >10 thousand of these used just once
- ~18 thousand used five or fewer times
- ~250 addresses used 100+ times

28

Mapping addresses to names

- Many-to-one
- Results

29

Compromises

- Process
 - Cannot be completely automated
 - Cannot be done with complete accuracy
 - Cannot, as a practical matter, be completed
- Limited objectives
 - Identify owners of much-used addresses
 - Find other addresses (“aliases”) used by those individuals

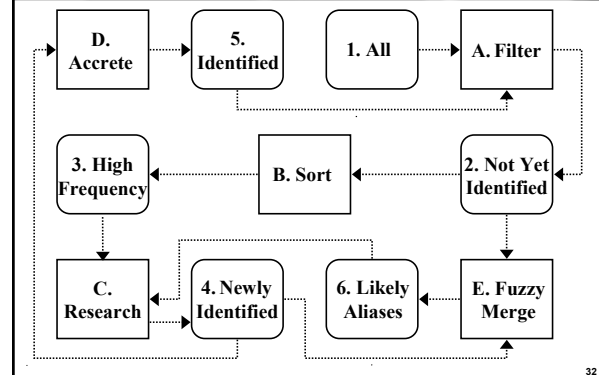
30

Process, in general

- Semi-automated research and analysis
- Reliant on triage
- Iterative; results of one cycle provide input for the next.
- Imperfect: false positives and negatives
- Manual aspects
 - Filter out false positives from mechanical matching
 - Use Google™ Groups to discover names

31

Flowchart



32

My paper about the mapping

- Paper 073-29: “Using Edit-Distance Functions to Identify “Similar” E-Mail Addresses”
- Coders' Corner
- Wednesday, May 12, 9:20 AM
- Room 519

33

About the Speaker

Howard Schreier

**U.S. Dept. of Commerce
Washington DC 20230
USA**

(202) 482-4180

**Howard_Schreier@ita.doc.gov
<http://howles.com/sugi/>**

34