

Laboratoire
d'Histoire des Sciences
et de Philosophie
Archives Henri Poincaré

Scientific mailing lists as a means of communication: A novel kind of corpus

Camille Paloque Bergès
Alexandre Hocquet

European Society for the History of Science Lisboa, Setembro 2014

ESHS

6th International Conference
of the European Society for the History of Science

Communicating
Science, Technology and Medicine

Lisboa
4-6 September 2014

Organised by
UIUHCT - FCUL/FCT-UNL

Outline

1. What is my story about? computational chemistry and software

2. What is my corpus? Computational Chemistry mailing List

3. Methodology: The flame wars

4. Methodology: The threaded conversation

5. Small data mining

Re: CCL:RTFM? What FM?

- **From:** "Peter Shenkin" <shenkin : at : still3.chem.columbia.edu>
- **Subject:** Re: CCL:RTFM? What FM?
- **Date:** Thu, 22 May 1997 15:29:27 -0400

I've gotten lots of private email responses on this subject, in addition to the posted responses I've seen. I've been archiving all of these.

After the dust settles, I'll post a summary to CCL.

-P.

--

**** "Deep Blue can't triumph in the game of life" (NY Times, 5/13/97)

* Peter S. Shenkin; Chemistry, Columbia U.; 3000 Broadway, Mail Code 3153 *

** NY, NY 10027; shenkin : at : columbia.edu; (212)854-5143; FAX: 678-9039

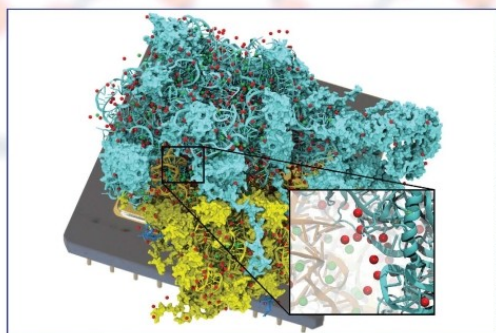
*MacroModel WWW page: <http://www.columbia.edu/cu/chemistry/mmod/mmod.html> *

1. Computational chemistry and software

JOURNAL OF Computational Chemistry

ORGANIC ■ INORGANIC ■ PHYSICAL ■ BIOLOGICAL

Volume 29 / Number 3 / February 2008
Articles Published Online in
Wiley InterScience,
25 June 2007 through 30 July 2007

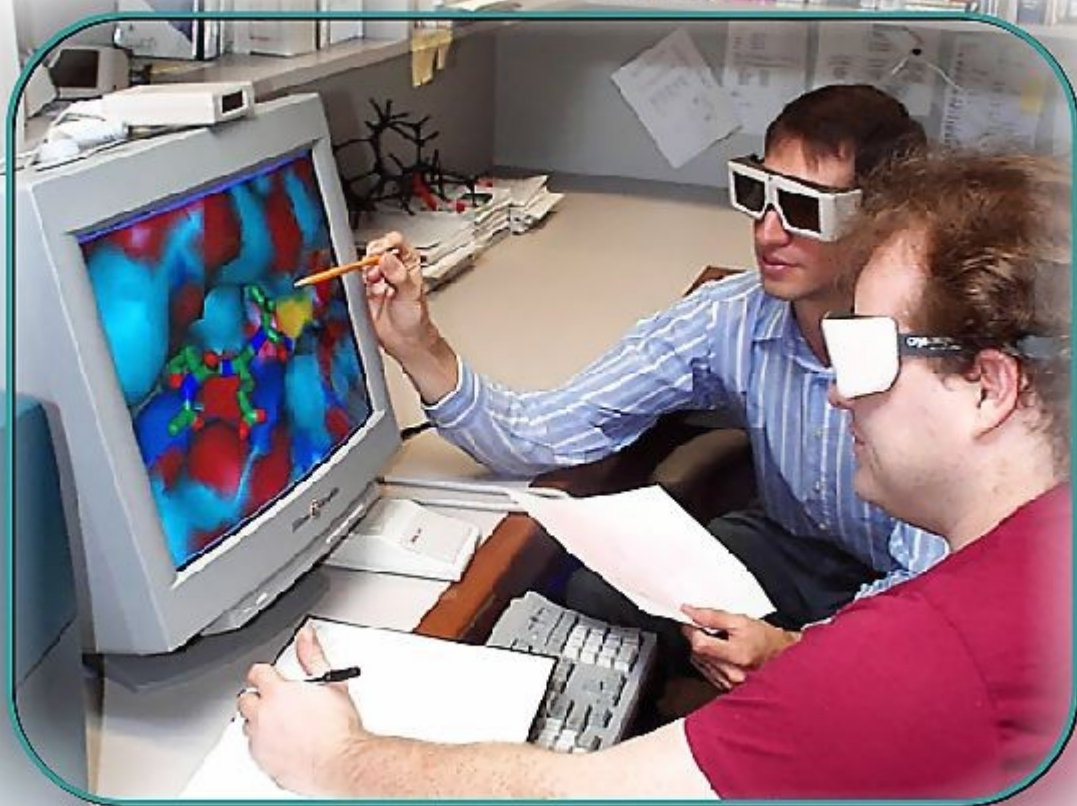


EDITORS: CHARLES L. BROOKS III • GERNOT FRENKING • SHIGEYOSHI SAKAI

WILEY
ISSN 0192-8651

ONLINE SUBMISSION AND PEER REVIEW
Wiley InterScience
www.interscience.wiley.com

Computational chemistry : use of computers to solve chemical problems



1. Computational chemistry and software



Computational chemistry & Hardware

- Supercomputers : 70s

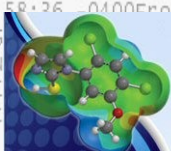
- Personal Computers : 90s

Computational chemistry & Software

- Developer & User



GaussView



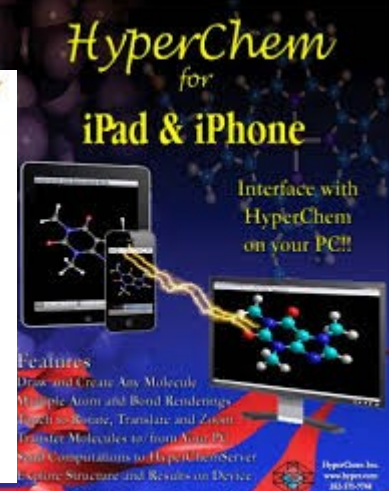
Spartan
Student Edition



Gaussian, Inc.
Carnegie Office Park - Building 6
Pittsburgh PA 15106 USA



NWChem
HIGH-PERFORMANCE COMPUTATIONAL
CHEMISTRY SOFTWARE



HyperChem
for
iPad & iPhone

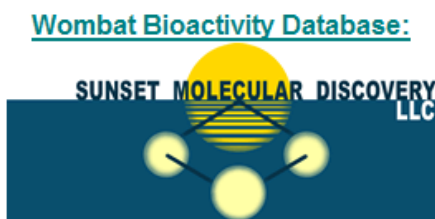
Interface with
HyperChem
on your PC!!

Features
Draw and Create Any Molecule
Simple Atom and Bond Renderings
Graph Rotate, Translate and Zoom
Transfer Molecules to/from Any PC
Local Computations with Public Server
Export Structure and Results on Device

2. The Computational Chemistry mailing List as a corpus

CCL: Computational Chemistry List

CCL.NET
computational chemistry list, ltd.



<http://www.ccl.net/index.shtml>

CCL Home

Messages in a queue: 0 (approx. wait time: 00:00 hours)

Page supported by:



[Computational Science & Engineering](#)
... a laboratory without walls for science

Welcome to the CCL Website!

ACS-COMP Division supports CCL:

semiempirical calculations simple oxidesFrom: sender@chemdc1.tau.ac.il (sender) Subject: Expert Know-How wanted !!Date: 02 Mar 1993 08:13:18 -0500 (EST) From: Gene Carter (Gene Carter) Subject: Co MD parameters.Date: Tue, 2 Mar 93 17:22:11 From: ...
lling Software: Does anyone have STELLA?Date: 04 Mar 1993 17:08:31 -0500 (EST) From: ...
ar 93 17:30:16 GMTDate: Sun, 7 Mar 93 12:36 ESTFrom: <SML108@PSUVM.PSU> Subject: MO Theory Made VisibleDate: Tue Mar 09 11:23:12 1993
uplingDate: Thu, 11 Mar 93 18:54 EDTFrom: <PODOSNNA%NYUACF1.BITNET@OHIO-STATE.EDU> Subject: partial charges for fictive atoms in G92Date: 15 Mar 1993 21:55:05 -0600 (CST) From: Chris Parkinson <mbdts@hpc.ch.mil>
ar 1993 13:58:36 -0400From: ZSHI@ac.dal.ca Subject: Information about IBM Girona Seminar on Molecular SimilarityDate: Thu, 18 Mar 93 10:11 -0500 Subject: Keep the noise down, pleaseDate: 19 Mar 93 11:37:00 EST
23 -0600 (CST) From: Andy Holder <AHOLDER@VAX1.UMKC.EDU> Subject: Announcing what about simulated annealing?Date: wed, 24 Mar 93 14:11:32
ject: SC&A Conference speakers - thanksDate: Thu, 25 Mar 93 10:18:48 -0500 From: MO@CHEMNA.DICHI.UNINA.ITS Subject: looking for ACES programFrom: Robert Laiuk
uk> Subject: Polyrate queryDate: Tue, 30 Mar 1993 19:31:32 +0200 From: c62@ulz) Subject: Stephen R. WilsonDate: Thu, 1 Apr 1993 16:33:29 +0200 From: o

Attack of the killer workstations

- From: [jle #@ world.std.com](mailto:jle#@world.std.com) (Joe M Leonard)
- Subject: Attack of the killer workstations
- Date: Tue, 10 Dec 91 20:04:44 -0500

Reviewing the recent discussion regarding C vs. Fortran and structured vs. spaghetti coding, isn't it reasonable to consider the time of the programmer as more important than the code's runtime (within reason)? Given that approx. \$20k workstations are comparable to Cray machines, isn't 10-ish% overhead more than worth the human time savings through better-thought-out programming (structured, modular, OO, etc.)? I hope the day's past for really needing to count cycles in loops - as one can wait for the next version of workstation xxx to gain appreciable speedups. Obviously, this isn't the case for mpp machines, for which the consensus is that one must rethink and/or reimplement algorithm's in order to use the hardware to advantage. Still, this might prove to be an example where clearer coding styles might turn out to be a win after all...

Joe Leonard

[jle #@ world.std.com](mailto:jle#@world.std.com)

2. The Computational Chemistry mailing List as a corpus

CCL.NET

computational chemistry list, ltd.

Send Message to CCL Subscribers or Administrator

For computer security reasons you have to retype this number: **567** in this box: . Your submission will not be accepted without

Please fill in all the fields below. Note: if you are sending a message to CCL Subscribers you will be asked for confirmation. After you click

URL in the confirmation e-mail or cut/paste the URL into your browser. This is the only way to verify that you entered your e-mail address

would be able to enter your e-mail address in the message, and impersonate you.

Click on the button when finished.

Important message to participants with Yahoo email addresses. [\[Please click here \]](#)

First/Given Name Middle Name Last/Sur/Family Name

full disclosure of methods?

- From: "DR. DOUGLAS A. SMITH, UNIVERSITY OF TOLEDO" <DSMITH -AatT- uoft02.ut>
- Subject: full disclosure of methods?
- Date: Sat, 26 Jun 1993 11:21:56 -0500 (EST)

Mark Thompson recently wrote, regarding the SAM1 topic currently under discussion:

>Let me address the more fundamental issue that this topic
>brings forth. I share Graham Hurst's concerns. One of the
>basic tenets of good science is that of reproducibility,
>and independant verification.

This is, I think, universally true and accepted. However, it is rarely followed. For example, there has been talk over the years that people who use molecular mechanics for their research should publish the parameters used for each study as part of the paper (or at least they should publish the differences between their parameters and the "standard" parameters

Administrator
Program to Ev...
sender@chemdc...
02 Mar 1993 08...
ers.Date: Tue,
e: 04 Mar 1993...
:36 ESTFrom: <S...
eDate: Tue Mar...
n: <PODOSNNA%NY...
<tom@cassandr...
ges for fictive...
is parkinson <ml...
Subject: Informa...
ce: Thu...
eDate: 19 Mar...
K1.UMKC.EDU>Sub...
?Date: wed, 24 f...
t: Thu, 25 Mar 93 10:18:48 -0500From: au195@cleveland.freenet.edu (Dr. Gene A. Nelson)Subject: Batching files in MOPACFrom: zrf10128@aw...
ACES programFrom: Robert Laird <mbdtsr1@hpb.ch.man.ac.uk>Subject: Fe-O calculationDate: Mon, 29 Mar 93 15:49:03 BSTFrom: rafapa@obel...
1993 19:31:32 +0200From: c62@aixfile1.urz.uni-heidelberg.de (Andreas Schulz)Subject: NBO analysesDate: Tue, 30 Mar 1993 16:04:42 -...
Apr 1993 16:33:29 +0200From: oles@gollum.uio.noSubject: BiosynDate: 01 Apr 1993 11:43:13 -0600 (CST)From: CUNDRIT@memstvx1.memst.edu

11:18:26 -0500From: jle@world.std.com (Joe...
CS.OHIO-STATE.EDUSubject: Steric strain at T...
IF calc's on diradicals/TSDDate: Mon, 18 Jan 9...
1:36 ESTFrom: landman@hal.physics.wayne.edu...
ct: Re: density problemDate: wed, 20 Jan 93...
'50:00 ESTFrom: nauss@wrair-emh1.army.milSU...
ussian90Date: 22 Jan 93 13:06:32 UTFFrom: "BD...
'31:46 -0500From: system@alchemy.chem.utoron...
m_ramsay@emunix.emich.eduSubject: CHEMICALC...
Jan 93 12:57:08 -0500From: watanabe@tammy.ha...
isted Chemistry Course Info RequestedDate: 27...
ct: Re: Periodic Table WidgetDate: wed, 27...
Population Analysis in GaussFrom: mei@ve...
e: Fri, 29 Jan 1993 16:47:38 -0600From: sliu...
an and excited statesDate: Fri, 29 Jan 93 15...
mentDate: 01 Feb 93 19:26:49-0500From: SBPM...
lect: DMF dielectricDate: Mon, 1 Feb 93 17:57...
ure -> 2D drawing?From: QINGSONG@minmet.lan...
software for calculating zeta potential and e...
Kevin Moore <kmoore@ncsc.org>Subject: MOLSOL...
'6:17:32 +0100From: ivan@gandalf.ciam.unibo.i...
emson.edu>Subject: Memory and CPU for Starde...
COM (Adi Treasurywala)Subject: Coordinates o...
allel workstations and serversDate: wed, 10...
-0800From: nicklahinsvm.comSubject: DIBUG ad...
resa windus)S...
&@OHSTVMA.ACS...
VERT@biolan.u...
3 11:29:01 PS...
DOUGLAS A. S...
(Cindy Fische...
hanker <ravi...
Feb 93 16:49...
L6A" <alex@ba...
MacDate: Thu...
sDate: Thu, 7...
guage questi...
993 15:25:19...
stryDate: Tue...
Tue, 2 Mar 9...
l@ccl.net>Dat...
erowitz Hanoc...
Subject: GRO...
n@quanta.phy...
unibas.ch (Re...
national freq...
2 code for UN...
ect: Software...
the ISTCP (pa...
8 ESTFrom: Ge...
Subject: Semic...
Christopher J

The Structural Transformation of the Public Sphere

An Inquiry into a Category of Bourgeois Society

Jürgen Habermas
translated by Thomas Burger
with the assistance of Frederick Lawrence

2. The Computational Chemistry mailing List as a corpus

Rules

By posting to the list, you [agree](#) to adhere to the rules outlined in this document.

These rules, regarding what topics are [acceptable/unacceptable](#) and [commercial announcements](#) discussions. Do not reopen Pandora's box -- please.

Acceptable Topics

Topics which **may appear on the list** include:

- **software/hardware-related** issues
 - bugs: Reporting bugs in chemistry software and possible work-arounds
 - programs and utilities
 - availability: Information on the availability of particular software, data, etc.
 - hardware-related issues relevant to the computational chemistry community
- **announcements** of
 - new chemistry software
 - computational chemistry-related workshops and symposia
- **methods**
 - new methods and techniques in computational chemistry
 - Q&A: questions and answers about solving computational chemistry problems
- **opinions** about services and products

Commercial Announcements

Commercial software: software sold to anyone for to the **list**, please respect the following rules:

- **Be brief** -- Short (25 lines or less) commercial p
 - **Be straightforward** -- The subject line should be
 - **No reposting** -- Reposts of software announcem
 - **No "vaporware"** -- Announcements of "vaporwa
- are not allowed.

The Structural Transformation of the Public Sphere

An Inquiry into a Category of Bourgeois Society

Jürgen Habermas

translated by Thomas Burger
with the assistance
of Frederick Lawrence

2. The Computational Chemistry mailing List as a corpus

IEEE
Annals

A Social History of Bitnet and Listserv, 1985-1991

April-June 2000 (vol. 22 no. 2)

pp. 32-41

David Alan Grier
Mary Campbell

DOI Bookmark: <http://doi.ieeecomputersociety.org/10.1109/85.841135>

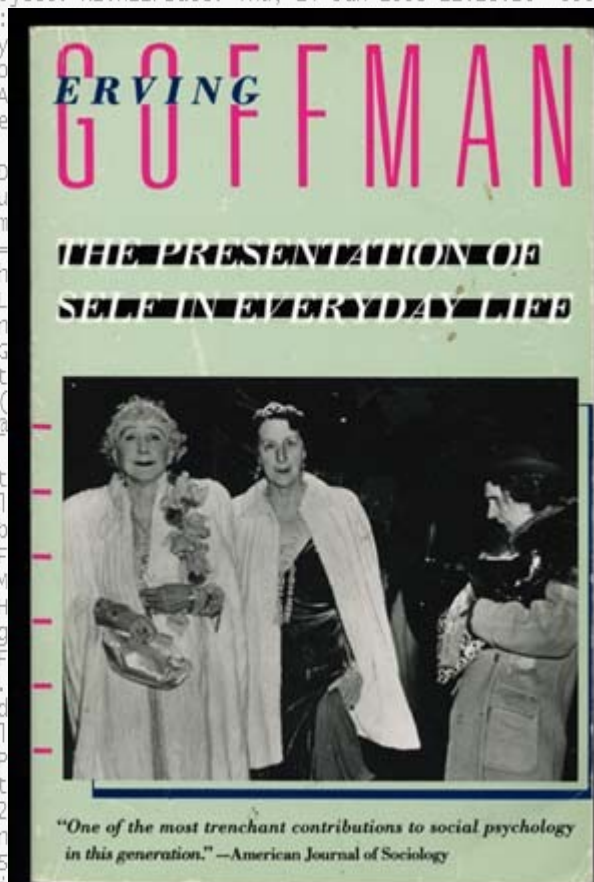
ABSTRACT

After the engineers built the computer networks, users had to build the social networks that made them useful. Listserv, the combined mailing list and file server, was an important tool for those interested in building network-based organizations. It first appeared on Bitnet, an academic network based on IBM computers. The early versions of Listserv became operational in the mid 1980s, and its early archives show how network users learned to use the software and, perhaps more importantly, how to manage network-based organizations.

Special Issue Articles

The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online

Bernie Hogan¹



Bulletin of Science, Technology & Society
30(6) 377-386

© 2010 SAGE Publications

Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>

DOI: 10.1177/0270467610385893

<http://bsts.sagepub.com>

SAGE

3. The flame wars

full disclosure of methods?

- From: "DR. DOUGLAS A. SMITH, UNIVERSITY OF TOLEDO" <[DSMTH -AatT- uoft02.ut](mailto:DSMTH-AatT-uoft02.ut)>
- Subject: full disclosure of methods?
- Date: Sat, 26 Jun 1993 11:21:56 -0500 (EST)

Re: full disclosure of methods?

Mark Thom
discussic
>Let me e
>brings i
>basic te
>and ind

- From: st-amant{*at*}cgl.ucsf.edu (Alain St-Amant)
- Subject: Re: full disclosure of methods?
- Date: 1993 12:21:56 -0700

Re: full disclosure of methods?

st-amant{*at*}cgl.ucsf.edu (Alain St-Amant)
disclosure of methods?
1993 12:21:56 -0700

wrote in the current discussion on the disclosure

did we had to reveal all our secrets and make them
d accessible? When software copyrights and patents
ate protection, maybe I will agree with that attitude.
smith is referring to the specific algorithms that are
the program more efficient but do not affect the final
e, I might agree.
disagree more if he is referring to some development
t actually affects the final results in any way.
ts me however, is the question of software copyrights
r. Smith alludes. I have been trying to get a feel
ghted and patented and I get a different answer from
eficic code be copyrighted or can the structure and
ted as well? How 'modified' should code be before
ly (and ethically if anyone is interested in expressing
ram? Or is it simply forbidden for code to

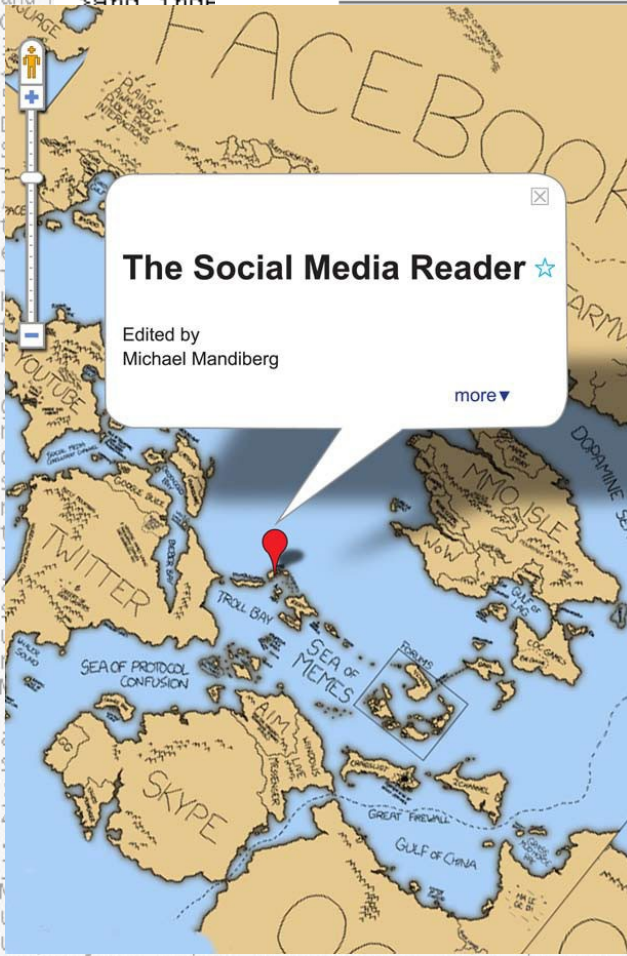
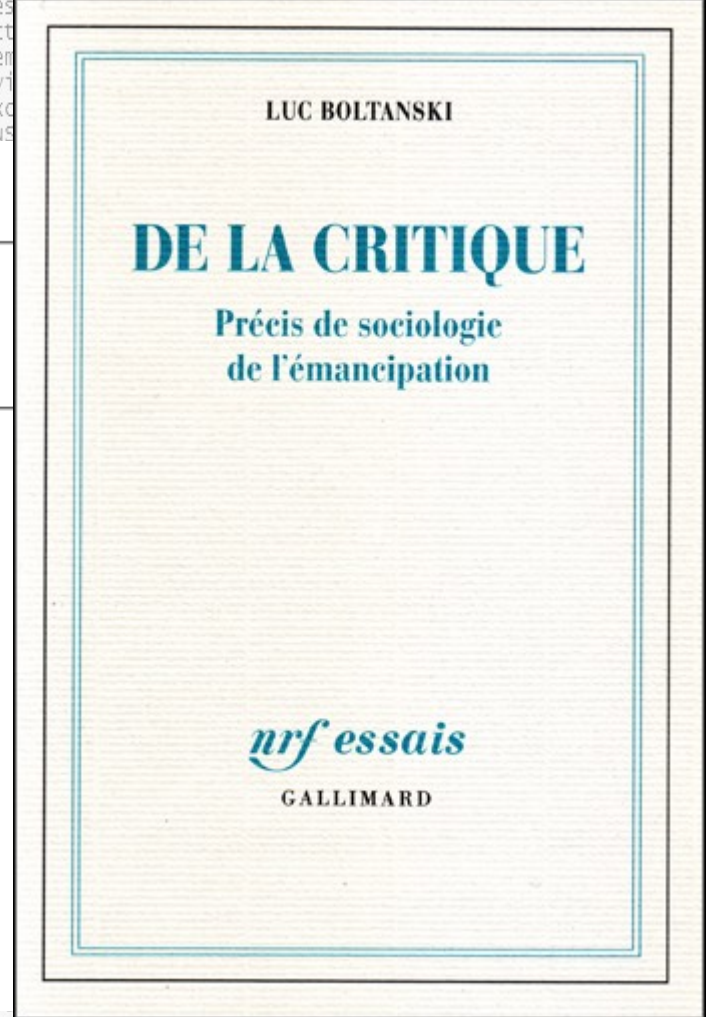
ne net any e-mail sent to me, but I think that this would
g discussion and it would be as interesting to hear how

Mar 93 11:37:00 ESTFrom: nauss@wrair-emh1.army.milSubject: Program called GRIDDate: Fri, 19 Mar 93 14:33:18 ESTFrom: Ge
US...@...Date: Fri, 19 Mar 93 17:26:57 -0800From: <AHOLDER@VAX1.UMKC.EDU>Subject: Semic
12...@...Date: Fri, 19 Mar 93 17:26:57 -0800From: <AHOLDER@VAX1.UMKC.EDU>Subject: Semic
Mar 93 10:18:48 -0500From: au195@cleveland.freenet.edu (Dr. Gene A. Nelson)Subject: Batchng files in MOPACFrom: zrf10128@aw
gramFrom: Robert Laird <mbdtsr1@hpb.ch.man.ac.uk>Subject: Fe-0 calculationDate: Mon, 29 Mar 93 15:49:03 BSTFrom: rafapa@obel
1931-0200...@...Date: Tue, 30 Mar 1993 16:04:42 -
6...@...Date: 01 Apr 1993 11:43:13 -0600 (CST)From: CUNDARIT@memstvx1.memst.edu

-Religious wars:

Flaming, Trolling

Sociologie des épreuves



Hacker and Troler as Trickster

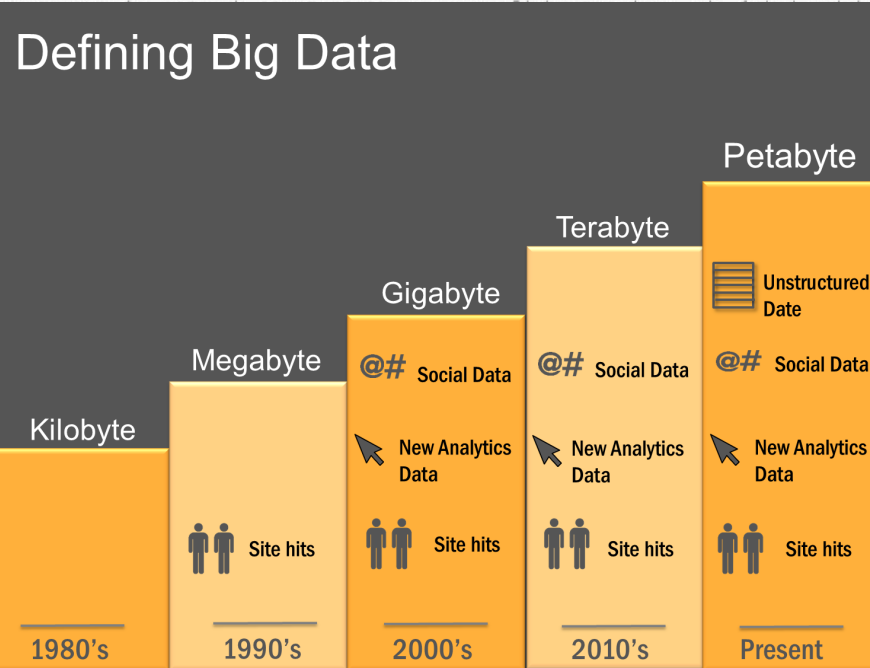
(Gabriella Coleman)

5. Small data mining

-a diachronic analysis of the debates : Time series

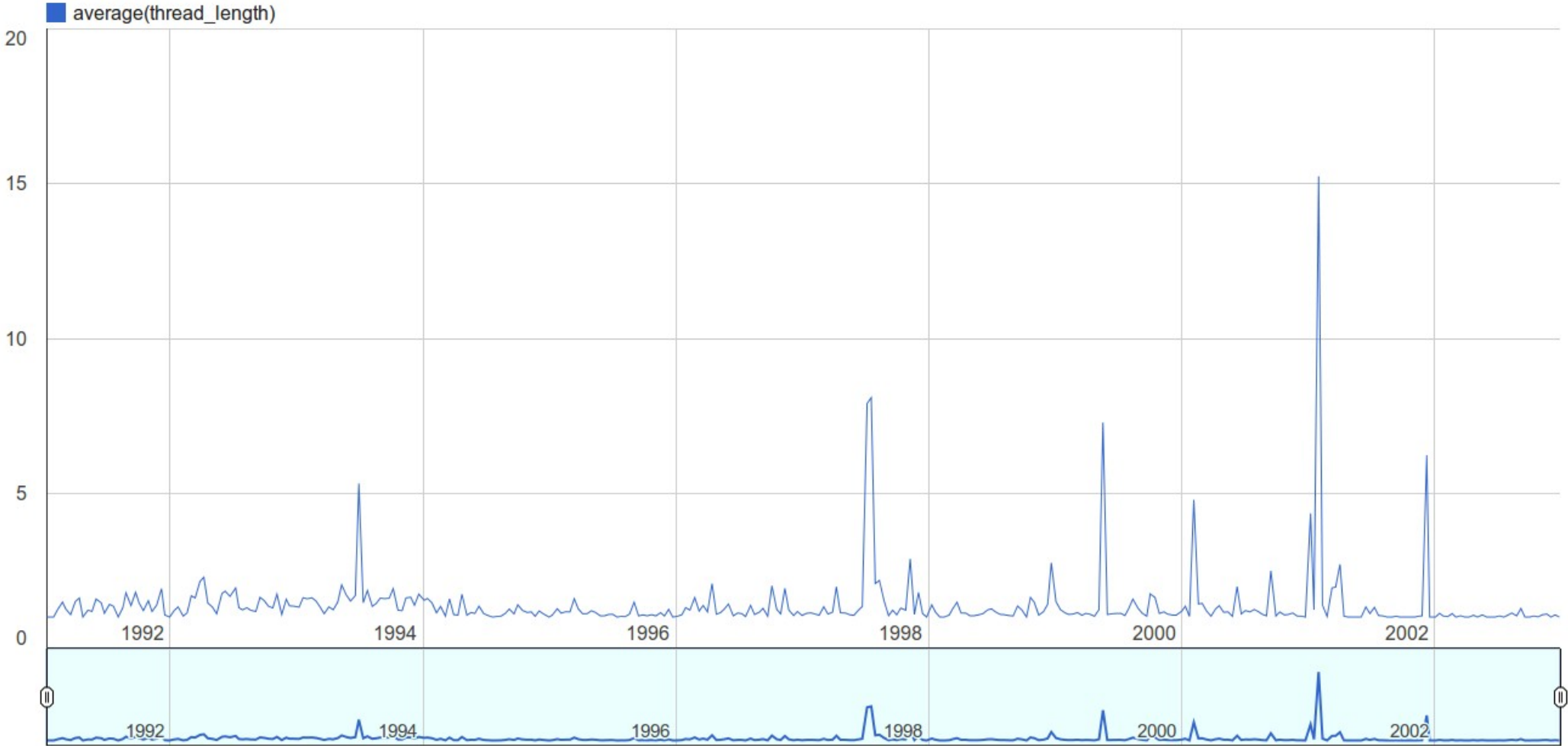
-a social network analysis : Bipartite graphs

-a literary analysis : Text mining



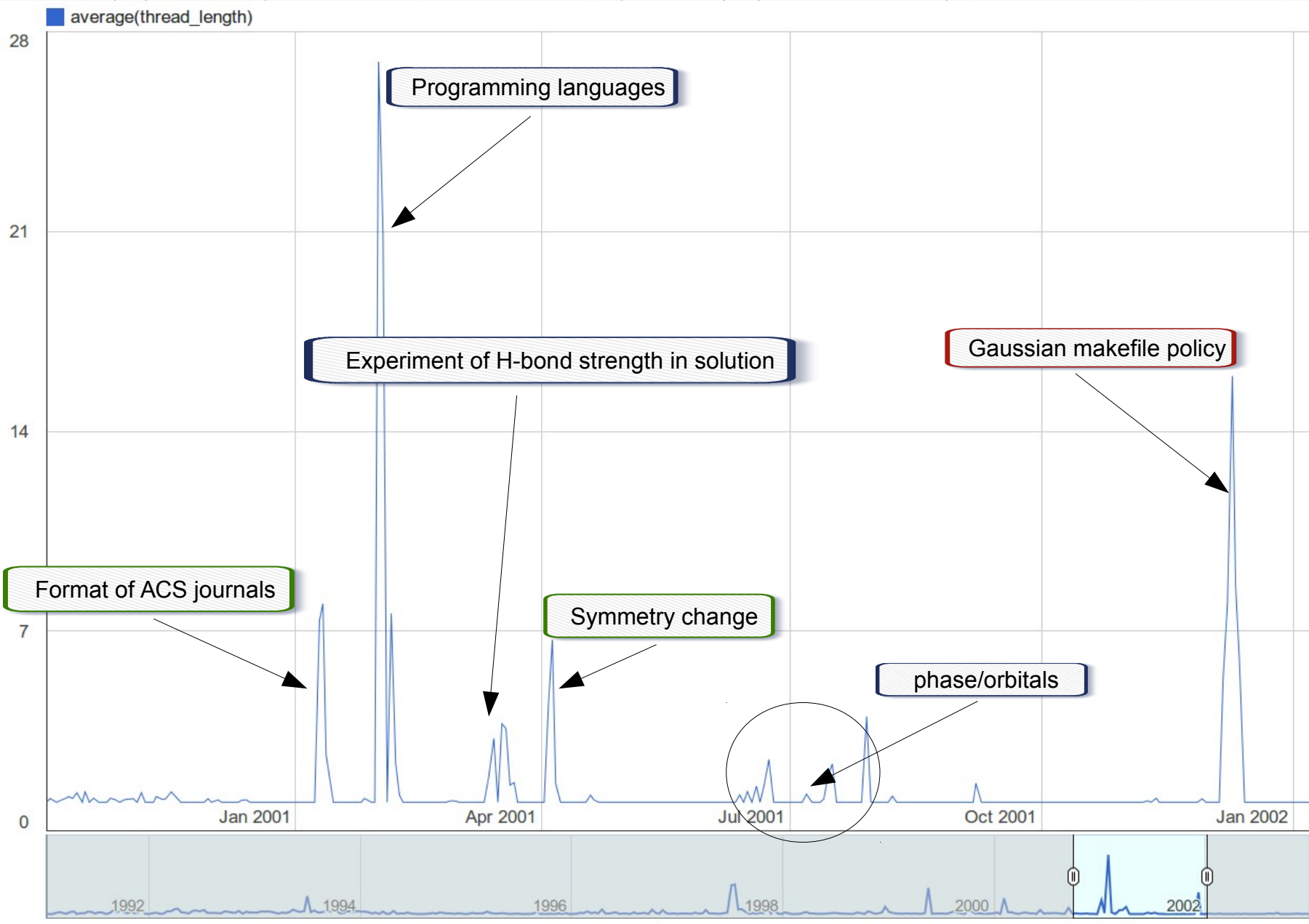
5. Small data mining: Time series

Averaged thread length vs. time: each peak represents an average of thread length for each message in a given period of time.
(if a message does not belong to a thread, it has a thread length of 1)



5. Small data mining: Time series

Year 2001 zoom: Six interesting threads extracted



Subject: Re: Taxol modellingDate: Fri, 1 Jan 93 14:11:05 ESTFrom: Francl Michelle M <mfranci@cc.brynmaur.edu>Date: Mon, 11 Jan 93 12:33:36 -0800From: hcj@gull.harvard.edu (Harry C. fzf-heidelberg.de>Subject: energy contribution of hbaonds?Date: 06 Jan 1993 10:55:29 -0500 (EST)From: "doug@lilly.com (Doug Johnson MC7R7)" <JOHNSON_DOUGLAS_W@LILLY.COM>Subject: u : CADD videosDate: Wed, 6 Jan 93 13:03:33 -0500From: sliu@chem.mcgill.ca (Shiuhong Li) <shihong@chem.mcgill.ca>Subject: PDB accessible surfacesDate: Thu, 7 Jan 93 !!Date: 08 Jan 1993 09:19:39 -0500 (EST)From: "doug@lilly.com (Doug Johnson MC7R7)" <JOHNSON_DOUGLAS_W@LILLY.COM>Subject: Replies to my posting about PDB searching methodsDate: F vi@tiberius.tc.cornell.edu>From: sscupham@reading.ac.ukDate: Mon, 11 Jan 93 17:54:05 GMTSubject: DMol questionsDate: Mon, 11 Jan 1993 14:33 ESTFrom: WILLIAMS%XRAY2@ulkyvxo2.loui or doubled messagesDate: Wed, 13 Jan 93 13:38:09 ESTFrom: rsefeck@watson.ibm.comSubject: data-flow programs for molecular modelingDate: Wed, 13 Jan 93 13:38:09 ESTFrom: rsefeck@ GAMESS and MOPACDate: Thu, 14 Jan 1993 10:42:35 -0500 (EST)From: HERMAN@ULNA.BWH.HARVARD.EDUSubject: RE:HELPPDate: Thu, 14 Jan 1993 11:18:26 -0500From: jle@world.std.com (Joe gov (Ron Shepard)Subject: Re: Cartesian coordinatesDate: Fri, 15 Jan 93 06:44 CDTFrom: GUNDA@MEMSTVMA.ACS.OHIO-STATE.EDUSubject: Steric strain at T 0From: dmschn@bbitt.monsanto.comSubject: Re: Molecular Graphics Society Meeting Date: Fri, 15 Jan 1993 15:00:00 ESTFrom: "JORG M. SEMINARIO" <JSMCM@royal.crc.uno.edu>Subject: Re: density problemDate: wed, 20 Jan 93 15:29:22 +0100From: oosten@ruihm1.LeidenUniv.nl (A.B. van Oosten)Subject: Re: proton+electron=neutron?Date: wed, 20 Jan 93 10:11:36 ESTFrom: landman@hal.physics.wayne.edu O-STATE.EDU>Subject: RE: PC Survey RESULTSDate: wed, 20 Jan 1993 15:47 CSTFrom: "JORGE M. SEMINARIO" <JSMCM@royal.crc.uno.edu>Subject: Re: density problemDate: wed, 20 Jan 93 ke to test a program.Date: Thu, 21 Jan 93 12:19:42 +0100From: Inst.f.Phys.Chem. I <pci@pctc.chemie.uni-erlangen.de>Date: 21 Jan 93 07:50:00 ESTFrom: nauss@wrair-emh1.army.milSub 93 13:06:32 UTFFrom: "BD tem@alchemistry.chem.utoron .eduSubject: CHEMICALC: From: watanabe@tammy.ha e Info RequestedDate: W le WidgetDate: wed, 27 in GaussianFrom: mei@ve 6:47:38 -0600From: sliu Date: Fri, 29 Jan 93 15 9:26:49-0500From: SBPM te: Mon, 1 Feb 93 17:57 n: QINGSONG@minmet.lan. ng zeta potential and e csc.org>Subject: MOLSOL an@gandalf.ciam.unibo.i nory and CPU for Starde)Subject: Coordinates o d serversDate: wed, 10 ym.comSubject: DIBUG ad o.gov (Theresa windus)S I-Linz.AC.%@OHSTVMA.ACS From: PREINERT@biolan.u , 16 Feb 93 11:29:01 PS From: "DR. DOUGLAS A. S cripps.edu (Cindy Fische n: G. Ravishanker <ravi : Mon, 22 Feb 93 16:49: dre MICU P16A" <alex@ba lator for MacDate: Thu, MS-windowsDate: Thu, 7 bject: language questio n, 1 Mar 1993 15:25:19 onal chemistryDate: Tue e MacDate: Tue, 2 Mar 9 rowski <jkl@cc1.net>Dat c.il (Senderowitz Hanoc gallion)Subject: GROW vTFrom: jim@quanta.phy. comp.bioz.unibas.ch (Re n with vibrational freq ct: Re: MM2 code for UN a.gov>Subject: Software ngress of the ISTCP (pa 93 14:33:18 ESTFrom: Ge JMCC.EDU>Subject: Semic 92From: "Christopher J MOPACFrom: zrf10128@aw 03 BSTFrom: rafapa@obel 30 Mar 1993 16:04:42 - RIT@memstvx1.memst.edu

5. Small data mining: Social Network Analysis

Subject: Re: Taxol modellingDate: Fri, 1 Jan 93 14:11:05 ESTFrom: Michelle M <mfranz1@cc.brynmawr.edu> Subject: u
 : CADD videosDate: Wed, 13 Jan 93 13:38:09 ESTFrom: rsefeck@watson.ibm.comSubject: data-flow programs
 !!Date: 08 Jan 1993 09:19:39 -0500 (EST)From: "dwj@lilly.com (Doug Johnson MC/R7)" <JOHNSON_DOUGLAS_W@LILLY.COM>Subject: replies to my posting about PDB searching methodsDate: F
 vi@tiberius.tc.cornell.edu>From: scsupham@reading.ac.ukDate: Mon, 11 Jan 93 17:54:05 GMTSubject: DMol questionsDate: Mon, 11 Jan 1993 14:33 ESTFrom: WILLIAMS%XRAY2@ulkyvxo2.loui
 or doubled messagesDate: wed, 13 Jan 93 13:38:09 ESTFrom: rsefeck@watson.ibm.comSubject: data-flow programs
 GAMES and MOPACDate: Thu, 14 Jan 1993 10:42:35 -0500 (EST)From: HERMAN@ULNA.BWH.HARVARD.EDUSubject: RE: EL
 gov (Ron Shepard)Subject: Re: Cartesian vs Internal coordinatesDate: Fri, 15 Jan 93 06:48 CDTFrom: CUNDARIT@MEMSTVX1.BITNET@OHSTVMA.ACS.OHIO-STATE.EDUSubject: Steric strain at T
 OFrom: dmschn@bbt.monsanto.comSubject: next Molecular Graphics Society MeetingFrom: m.e. <harbowy@chemres.tn.cornell.edu>Subject: UHF calc's on diradicals/TSDDate: Mon, 18 Jan 9
 n 93 15:29:22 +0100From: oosten@ruihl1.LeidenUniv.nl (A.B. van Oosten)Subject: Re: proton+electron+neutron
 O-STATE.EDU>Subject: RE: PC Survey RESULTSDate: Wed, 20 Jan 1993 15:47 CSTFrom: "JORGE M. SEMINARIO" <JSMC@rcival.crc.unn.edu>Subject: Re: density problemDate: wed, 20 Jan 93
 ke to test a program.Date: Thu, 21 Jan 93 12:19:42 +0100From: Inst.f.Phys.Chem. I <pci@pctc.chemie.uni-erlangen.de>Subject: Re: density problemDate: wed, 20 Jan 93

Filtered 2001 network:

Green dots are email addresses

(size = # of messages)

Red dots are threads/topics

(size = # of contributors)

Edges link red and green dots

when a contributor post in a topic

(weighted by # of messages)

Programming languages

Format of ACS journals

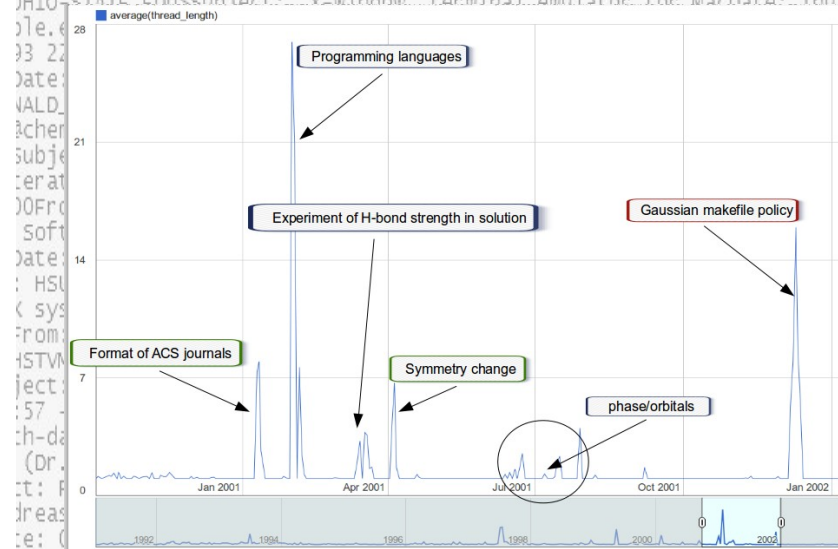
Gaussian policy

« symmetry change » is isolated:

It does not attract the same contributors

Argument: large edges with few people

Debate: numerous people, small edges



5. Small data mining: Text mining

Summary

- There are 12 documents in this corpus with a total of 5,657,131 words and 238,437 unique words.
- Longest documents (by words): CCL1995.csv (610,247), CCL1996.csv (597,253). Shortest documents: CCL1991.csv (124,261), CCL1992.csv (233,877). All...
- Highest vocabulary density (): CCL1991.csv (117.8), CCL1992.csv (98.9). Lowest density: CCL2002.csv (73.0), CCL2000.csv (74.6). All...
- Most frequent words in the corpus: chemistry (17,534), dear (11,031), molecular (10,969), program (10,897), thanks (10,297). More...
- Words with notable peaks in frequency across the corpus: met (1,798), types (786), linux (3,448), sci (1,592), protein (3,102). More...
- Distinctive words (compared to the rest of the corpus)
 1. CCL1991.csv: chemistry (621), list (311), ftp (236), fortran (227), code (238). More...
 2. CCL1992.csv: ftp (337), mopac (385), list (418), chemistry (423), send (271). More...
 3. CCL1993.csv: program (1,157), netters (608), cgi (497), ftp (451), ibm (307). More...
 4. CCL1994.csv: netters (823), chemistry (1,950), g92 (408), program (1,256), ida-1.4.4 (205). More...
 5. CCL1995.csv: netters (694), program (1,407), methods (690), g92 (295), molecular (1,358). More...
 6. CCL1996.csv: form (463), g94 (459), mopac (583), dear (1,371), summary (772). More...
 7. CCL1997.csv: 3d (1,015), g94 (568), dft (485), computational (843), molecular (1,311). More...
 8. CCL1998.csv: linux (501), g94 (391), dear (1,002), ft (144), nt (176). More...
 9. CCL1999.csv: g98 (532), xml (250), wrote (554), best (541), linux (486). More...
 10. CCL2000.csv: g98 (660), help (938), b0 (225), hi (747), cclers (379). More...
 11. CCL2001.csv: g98 (776), div (759), 3d (963), html (511), br (475). More...
 12. CCL2002.csv: 3d (1,537), div (1,176), html (702), size (691), face (562). More...

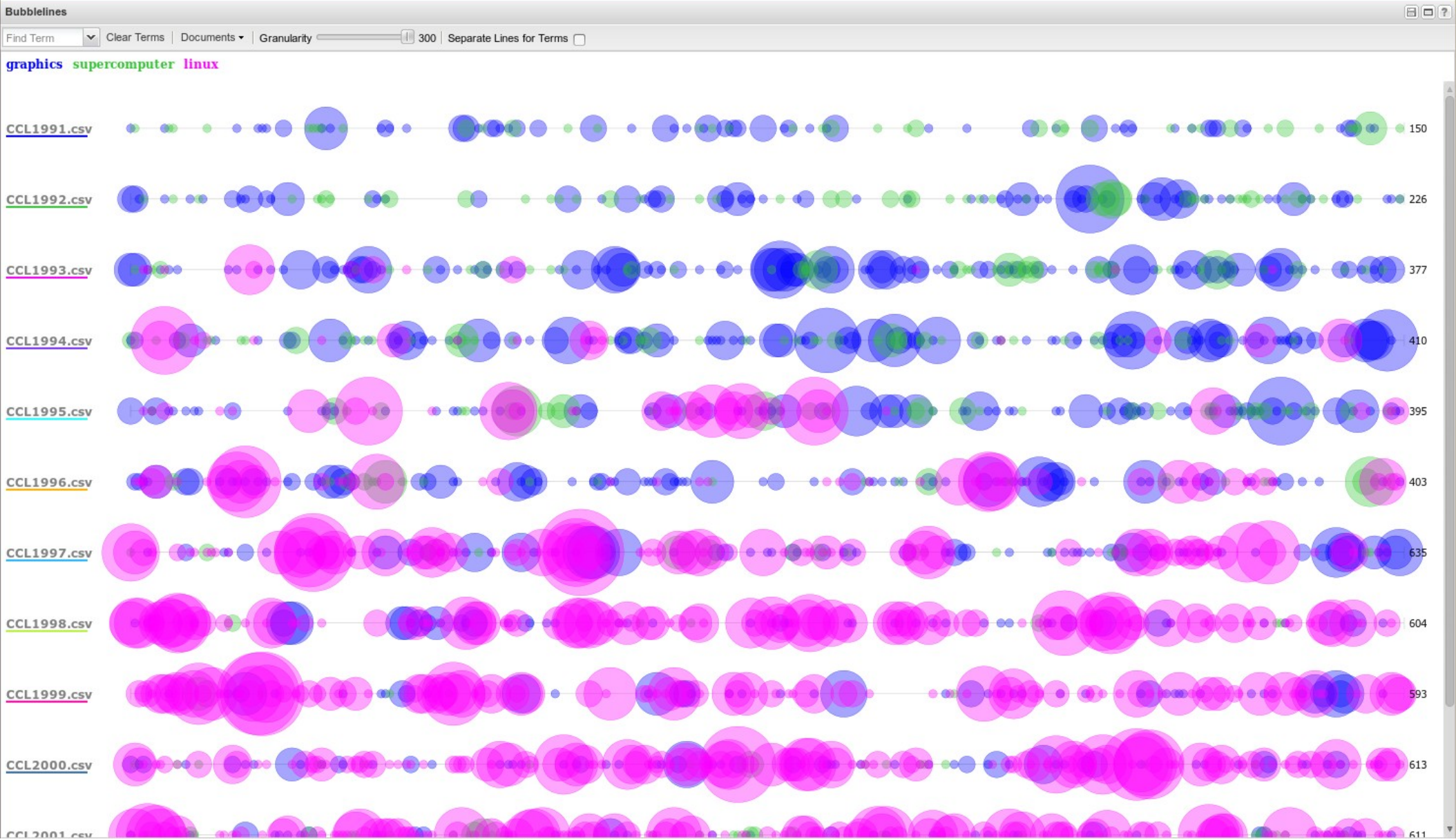
Notable peaks in frequency

Distinctive words compared to the rest of the corpus

5. Small data mining: Text mining

Periodizing the corpus according to the nature of the debates.

- « graphics » : blue
- « supercomputer » : green
- « linux » : pink



Subject: Re: Taxol modellingDate: Fri, 1 Jan 93 14:11:05 ESTFrom: Francl Michelle M <mfran1@cc.brynmaur.edu>Date: Mon, 4 Jan 93 12:33:36 -0800From: hcj@gull.uncc.edu (Harry C. kfz-heidelberg.de>Subject: energy contribution of hheads?Date: 06 Jan 1993 10:55:29 -0500 (EST)From: "doug@lilly.com (Doug Johnson MC7R7)" <JOHNSON_DOUGLAS_W@LILLY.COM>Subject: u : CADD videosDate: Wed, 6 Jan 93 13:03:33 -0500From: slm@vlsi.com (Srinivasan Madhav) <slm@vlsi.com>Subject: How to find accessible surfacesDate: Thu, 7 Jan 93 !!Date: 08 Jan 1993 09:19:39 -0500 (EST)From: "dwj@lilly.com (Doug Johnson MC7R7)" <JOHNSON_DOUGLAS_W@LILLY.COM>Subject: replies to my posting about PDB searching methodsDate: F vi@tiberius.tc.cornell.eduFrom: ccsupham@reading.ac.ukDate: Mon, 11 Jan 93 17:54:05 GMTSubject: DMol questionsDate: Mon, 11 Jan 1993 14:33 ESTFrom: WILLIAMS%XRAY2@ulkyvx02.loui or doubled <mailto:williams@reading.ac.uk>Subject: DMol questions for molecular modelingDate: wed, 13 Jan 93 13:38:09 ESTFrom: rsefeck@

5. Small data mining: Text mining

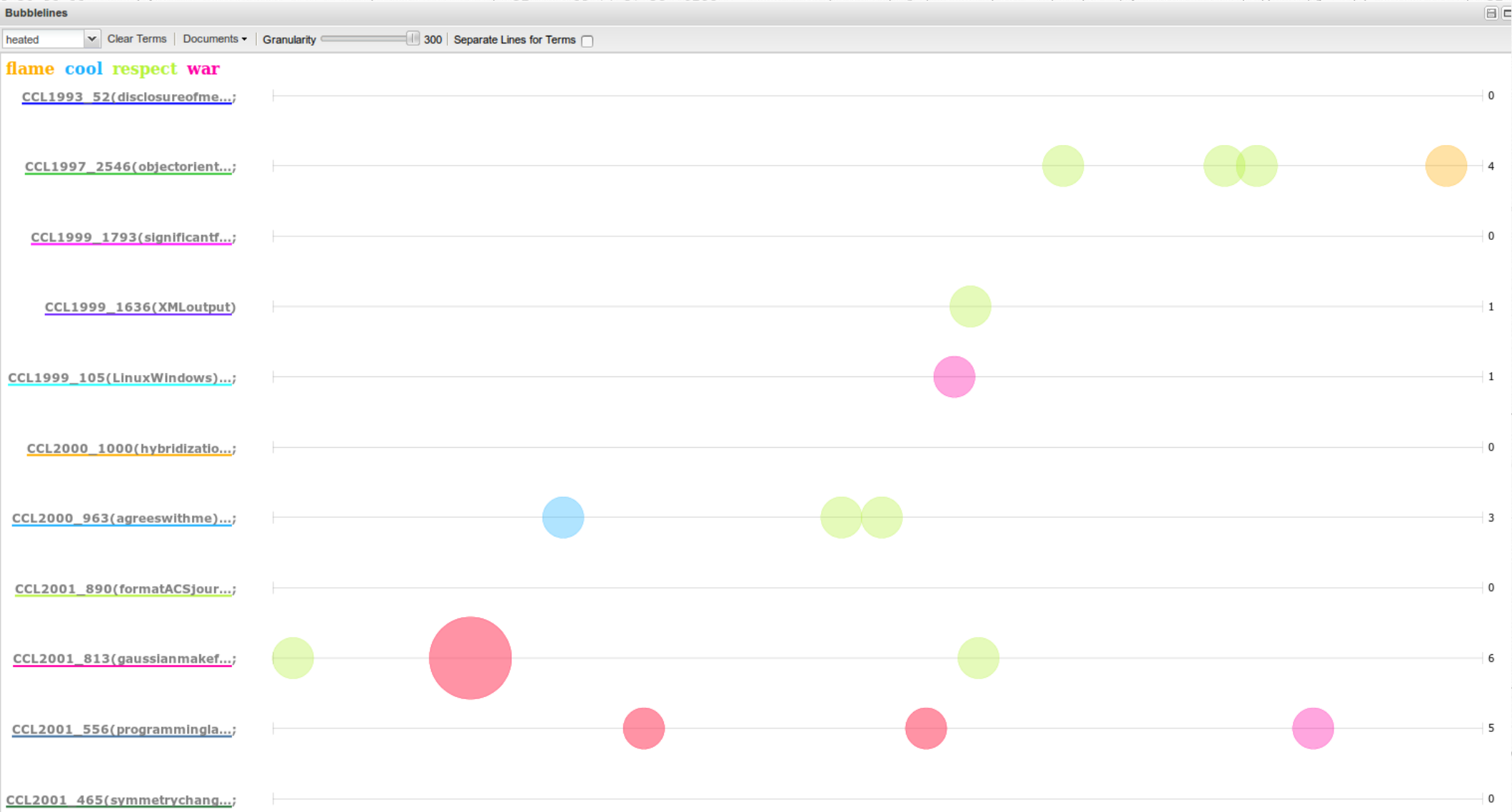
Another kind of partitioning: according to the threads



5. Small data mining: Text mining

Subject: Re: Taxol modellingDate: Fri, 1 Jan 93 14:11:05 ESTFrom: Francl Michelle M <mfrancl@cc.brynmawr.edu>Date: Mon, 4 Jan 93 12:33:36 -0800From: hcj@gull.uncc.edu (Harry C. kfz-heidelberg.de>Subject: energy contribution of hba...Date: 06 Jan 1993 10:55:29 -0500 (EST)From: "doug@lilly.com (Doug Johnson MC7R7)" <JOHNSON_DOUGLAS_w@LILLY.COM>Subject: u : CADD videosDate: Wed, 6 Jan 93 13:03:33 -0500From: sfinn@omni.cc.brown.edu (Steve Finney)Subject: re: molecular modeling accessible surfacesDate: Thu, 7 Jan 93 !!Date: 08 Jan 1993 09:19:39 -0500 (EST)From: dmj@lilly.com (Doug Johnson MC7R7) <JOHNSON_DOUGLAS_w@LILLY.COM>Subject: replies to my posting about PDB searching methodsDate: F vi@tiberius.tc.cornell.edu>From: scsupham@reading.ac.ukDate: Mon, 11 Jan 93 17:54:05 GMTSubject: DMol questionsDate: Mon, 11 Jan 1993 14:33 ESTFrom: WILLIAMS%XRAY2@ulkyvx02.loui or doubled messagesDate: wed, 13 Jan 93 13:38:09 ESTFrom: rsefeck@watson.ibm.comSubject: data-flow programs for molecular modelingDate: wed, 13 Jan 93 13:38:09 ESTFrom: rsefeck@ GAMESS and MOPACDate: Thu, 14 Jan 1993 10:42:35 -0500 (EST)From: HERMAN@ULNA.BWH.HARVARD.EDUSubject: RE:HELPPDate: Thu, 14 Jan 1993 11:18:26 -0500From: jle@world.std.com (Joe gov (Ron Shepard)Subject: Re: Cartesian vs Internal coordinatesDate: Fri, 15 Jan 93 06:48 CDTFrom: CUNDARIT%MEMSTVX1.BITNET@OHSTVMA.ACS.OHIO-STATE.EDUSubject: Steric strain at T 0From: dmschn@bbbit.monsanto.comSubject: next Molecular Graphics Society MeetingFrom: m.e. <harbowy@chemres.tn.cornell.edu>Subject: UHF calc's on diradicals/TSDate: Mon, 18 Jan 9 n 93 15:29:22 +0100From: oosten@ruihml.LeidenUniv.nl (A.B. van Oosten)Subject: Re: proton+electron=neutron?Date: wed, 20 Jan 93 10:11:36 ESTFrom: landman@hal.physics.wayne.edu O-STATE.EDUSubject: RE: PC Survey RESULTSDate: wed, 20 Jan 1993 15:47 CSTFrom: "JORGE M. SEMINARIO" <JSMCM@royal.crc.uno.edu>Subject: Re: density problemDate: wed, 20 Jan 93 ke to test a program.Date: Thu, 21 Jan 93 12:19:42 +0100From: Inst.f.Phys.Chem. I <pci@pctc.chemie.uni-erlangen.de>Date: 21 Jan 93 07:50:00 ESTFrom: nauss@wrair-emh1.army.milSub Jan 93 13:27:12 GMTFrom: "D. Mathieu" <DEWINDT%ERSACL1.BITNET@OHSTVMA.ACS.OHIO-STATE.EDUSubject: force constant symmetry and Gaussian90Date: 22 Jan 93 13:06:32 UTFFrom: "BD le Recurrence" <le@chemres.tn.cornell.edu>Subject: RE:: PDB on the MacDate: Thu, 21 Jan 93 17:31:46 -0500From: system@alchemy.chem.utoron From: arne@mango.mef.ki.se (Arne Elofsson)Subject: Re: charmm and SGIS - charmm newsgroup?Date: Sun, 24 Jan 93 15:05:30 -0500From: chm_ramsay@emunix.emich.eduSubject: CHEMICALC: Subject: This is testDate: 26 Jan 1993 07:20:59 -0600 (CST)From: AHOLDER@VAX1.UMKC.EDUSubject: semiempirical referencesDate: Tue, 26 Jan 93 12:57:08 -0500From: watanabe@tammy.ha ate: 27 Jan 93 08:59:32-0500From: /PN=DAVID.W.ELROD/DD.ID=SSW.DWELROD/PRMD=UPJOHN/ADMD=TELEMAIL/C=US/@sprint.comSubject: Computer Assisted Chemistry Course Info RequestedDate: W

Locating the most passionate threads



Bibliography

Boczkowski, P. J. 1999. "Mutual shaping of users and technologies in a national virtual community." *Journal of Communication* 49(2):86-108.

Bury, Rhiannon. 2005. *Cyberspaces Of Their Own: Female Fandoms Online*. Peter Lang.

Flichy, Patrice. 1999. "Internet ou la communauté scientifique idéale." *Réseaux* 17(97):77-120.

Grier, D.A., and M. Campbell. 2000. "A social history of Bitnet and Listserv, 1985-1991." *IEEE Annals of the History of Computing* 22:32-41. Retrieved October 17, 2011.

Hansen, Derek L., Ben Shneiderman, and Marc Smith. 2010. "Visualizing threaded conversation networks: mining message boards and email lists for actionable insights." Pp. 47-62 in *Proceedings of the 6th international conference on Active media technology, AMT'10*. Berlin, Heidelberg: Springer-Verlag

Kelty, Christopher M. 2008. *Two Bits: The Cultural Significance of Free Software*. Duke University Press.

Mandiberg, Michael. 2012. *The social media reader*. New York: New York University Press.

Marshall, Jonathan Paul. 2007. *Living on Cybermind: Categories, Communication, and Control*. Peter Lang.