complex aspects of seemingly simple information needs

jussi karlgren, gavagai and kth, stockholm

april 2017

jussi karlgren

- real scale real time distributional semantics

jussi karlgren

- real scale real time distributional semantics
- stylistics and genre in text and conversation

jussi karlgren

- real scale real time distributional semantics
- stylistics and genre in text and conversation
- since 2010 at the text analysis company Gavagai

jussi karlgren

- real scale real time distributional semantics
- stylistics and genre in text and conversation
- since 2010 at the text analysis company Gavagai
- and as adjoint professor of language technology at KTH

gavagai

- research-based lexical learning back end
  lexicon.gavagai.se

gavagai

- research-based lexical learning back end
  lexicon.gavagai.se
- media monitoring in >40 languages
  monitor.gavagai.se

gavagai

- research-based lexical learning back end
  lexicon.gavagai.se
- media monitoring in >40 languages
  monitor.gavagai.se
- highly interactive clustering and theme
  extraction from open answers to questionnaires
  etc
  explorer.gavagai.se

typical seemingly simple information needs

- What are customers and potential customers saying about my brand / about us / about some concept of interest on line?

# typical seemingly simple information needs

- What are customers and potential customers saying about my brand / about us / about some concept of interest on line?
- Are our customers happy?

typical seemingly simple information needs

- What are customers and potential customers saying about my brand / about us / about some concept of interest on line?
- Are our customers happy?
- What do our customers want from us?

typical seemingly simple information needs

- ► What are customers and potential customers saying about my brand / about us / about some concept of interest on line?
- ► Are our customers happy?
- ► What do our customers want from us?
- ► How could we make them happier?

typical seemingly simple information needs

- ▶ What are customers and potential customers saying about my brand / about us / about some concept of interest on line?
- ▶ Are our customers happy?
- ▶ What do our customers want from us?
- ▶ How could we make them happier?

easy to pose, easy to act on answers if results are delivered as expected
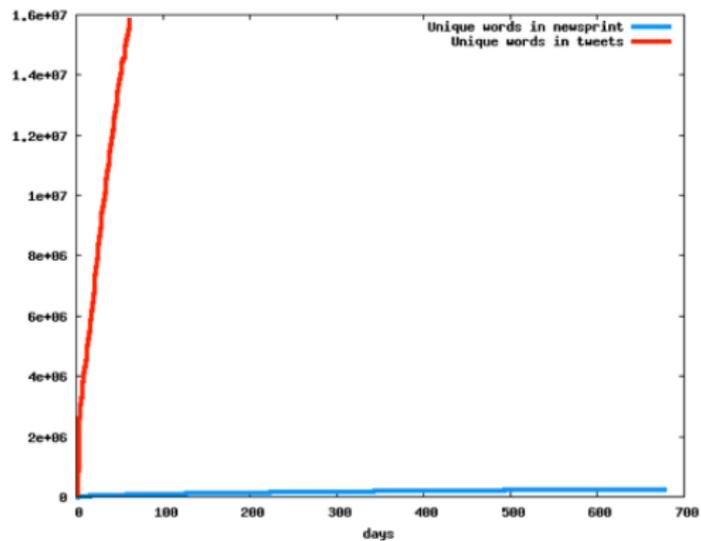
big data

big data and human information processing

- human information processing is really good at big data

big data and human information processing

- human information processing is really good at big data
- analogy
  & saliency

big data and human information processing

- human information processing is really good at big data
- analogy
  & saliency
- patterns and change rather than the literal

big data and human information processing

- human information processing is really good at big data
- analogy
  & saliency
- patterns and change rather than the literal
- self-learning rather than instruction

semantic base technology

- is this an example of that?
- are these two the same?
- has this changed? how?
- what is the relation of this and that?
- is this a new way of saying that?
- are these or those more like this?
- is this typical or strange?
- can we trust this?
- does the author believe this to be true?

semantic base technology

semantic base technology

| |
|---|
| Meaningful application |
| **Semantic layer** |
| Crunch layer |
| Distributed processing architecture |
| Database technology or similar |
| Data stream |

is this an example of that?
are these two the same?
has this changed? how?
what is the relation of this and that?
is this a new way of saying that?
are these or those more like this?
is this typical or strange?
can we trust this?
does the author believe this to be true?

learning a lexicon
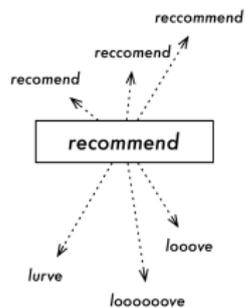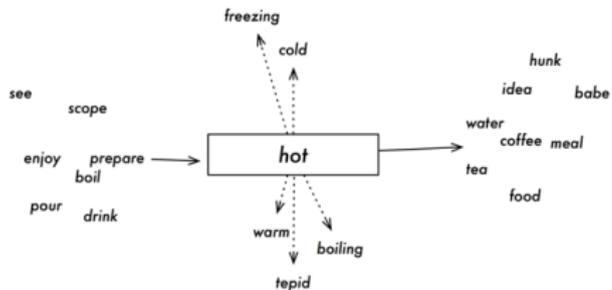
learning a lexicon

http://lexicon.gavagai.se
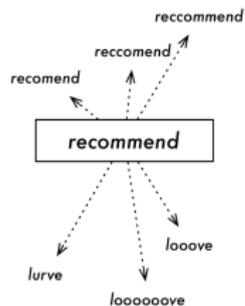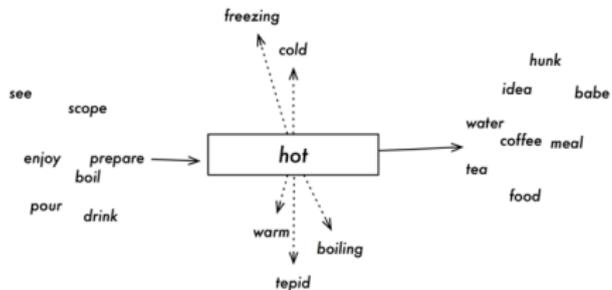
sentiment analysis

attitude is a non-conventionalised aspect of language use

sentiment analysis

attitude is a non-conventionalised aspect of
language use
... and therefore interesting as an object of study ...

sentiment analysis

attitude is a non-conventionalised aspect of
language use
... and therefore interesting as an object of study ...
... and potentially quite useful in commercial
application!

sentiment analysis can be difficult!

- And the sound quality - my God!
- Raymond left no room for error on his recordings and it shows.
- Definitely one of the better tracks on the album.
- Wow, could have been a expansion pack.
- I loved The Spy Who Came In From The Cold but the movie is a bit dated in a way the book never will be.
- Meat is more environmentally friendly than seafood.
- I am unsure about the feasibility of this knitting pattern.
- I love the Samsung B2710 but I would not recommend it to my colleagues.
- I don't know if I should call her up ? I liked her when I met her last weekend.
- This is true.

sentiment analysis
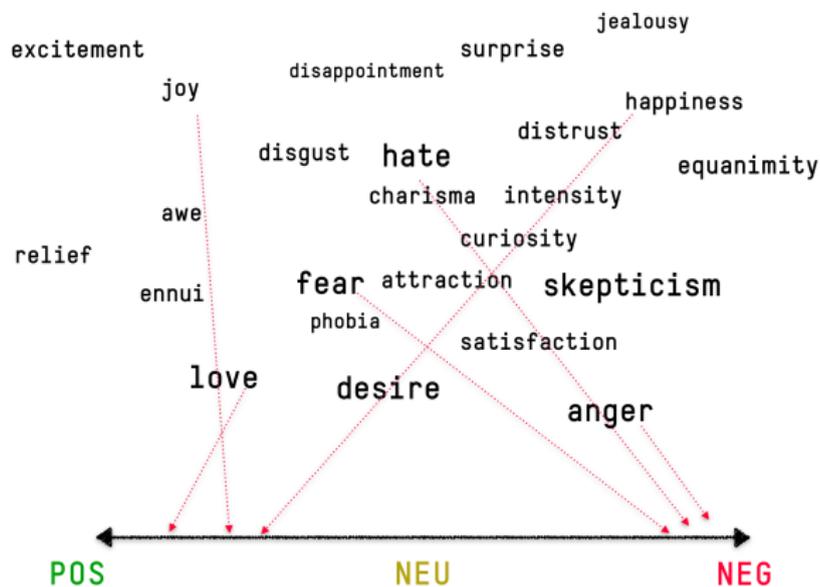
sentiment analysis

Possibly the most offensive misrepresentation I've seen

So @ChrisWarcraft is speaking my law class on Thursday.
I was excited, but after #GamerGate, I am now holy-
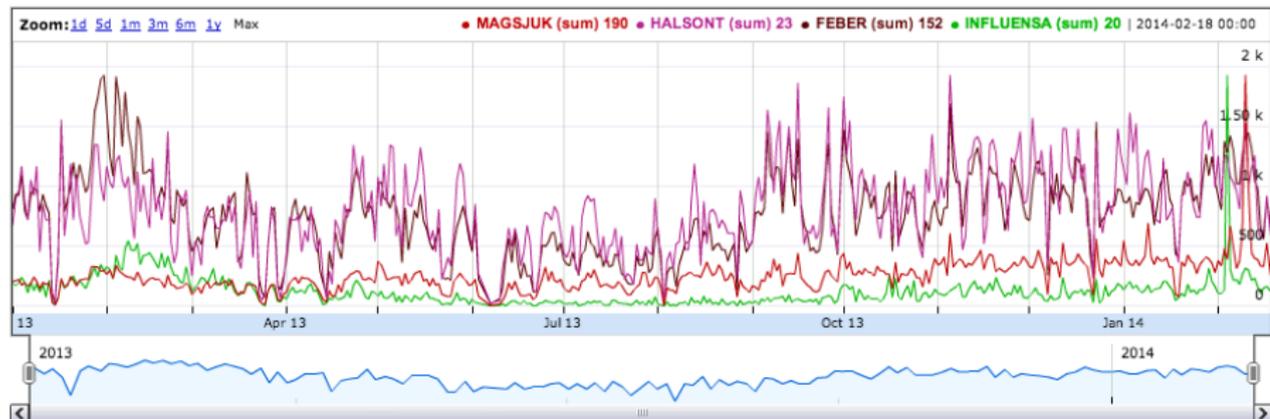mother-of-all-creation excited

**sentiment analysis? 10%
coverage wrt to POS-NEG**

piece on #Gamergate
o think up a whole

Oh man I love to done spiked armor with nipples and
fight for love and honor. Thanks for the compliment
#GamerGate

**multi-dimensional
tonality: > 50% coverage**

sentiment analysis

geolocation

predict author location from text: words carry information about position

geolocation

predict author location from text: words carry information about position

"I'm taking the tram now"

(few candidates)

geolocation

predict author location from text: words carry information about position

"I'm taking the tram now"

(few candidates)

"God I hate Stockholm, people are so stressed"

(inconclusive)

geolocation

predict author location from text: words carry information about position

"I'm taking the tram now"

(few candidates)

"God I hate Stockholm, people are so stressed"

(inconclusive)

"Oh lovely, lovely Falköping"

(fairly conclusive)

why do all this?

why do all this?
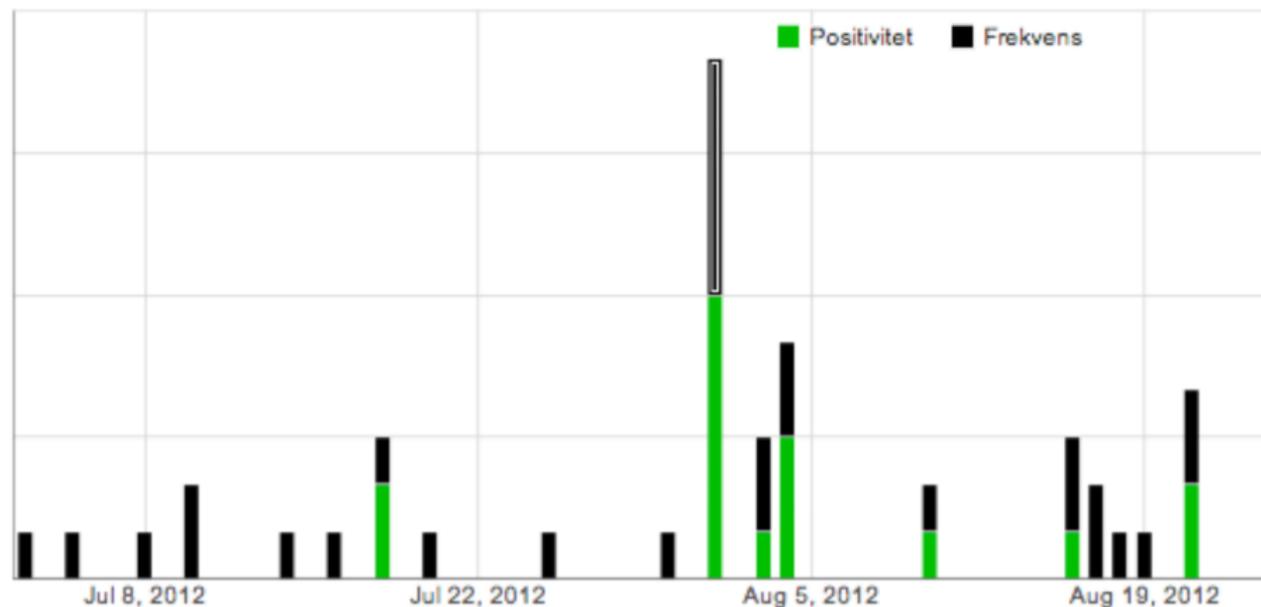
philological interest

why do all this?

philological interest

tools for market analysis, brand reputation mining, and other useful applications
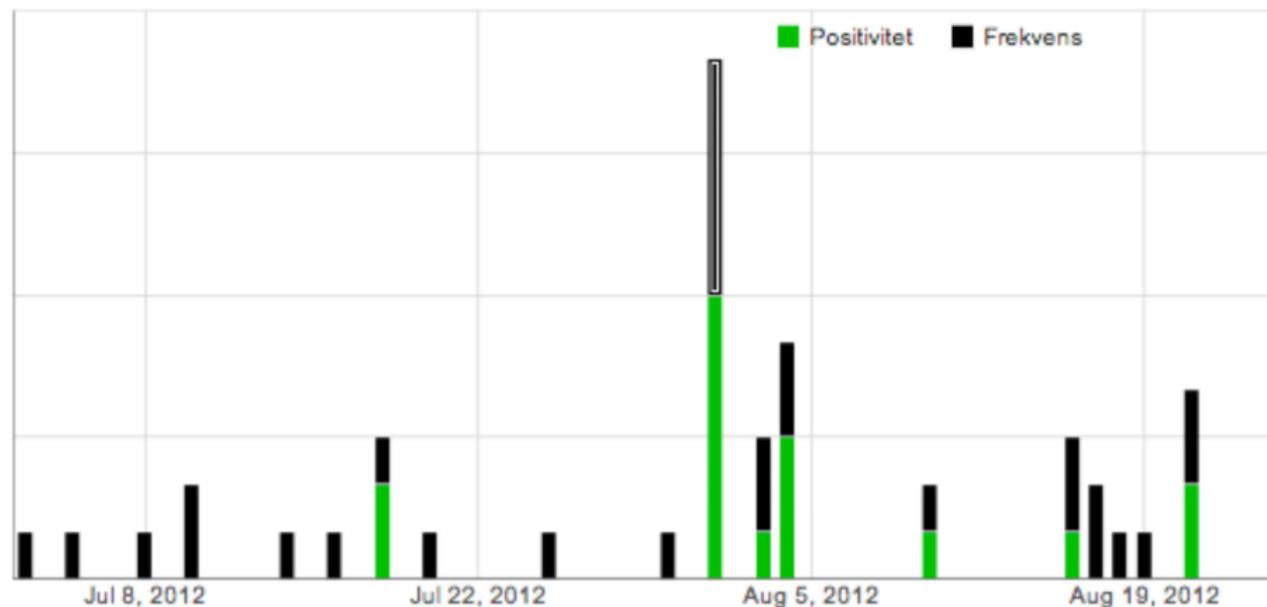
why do all this?

it is not all that simple
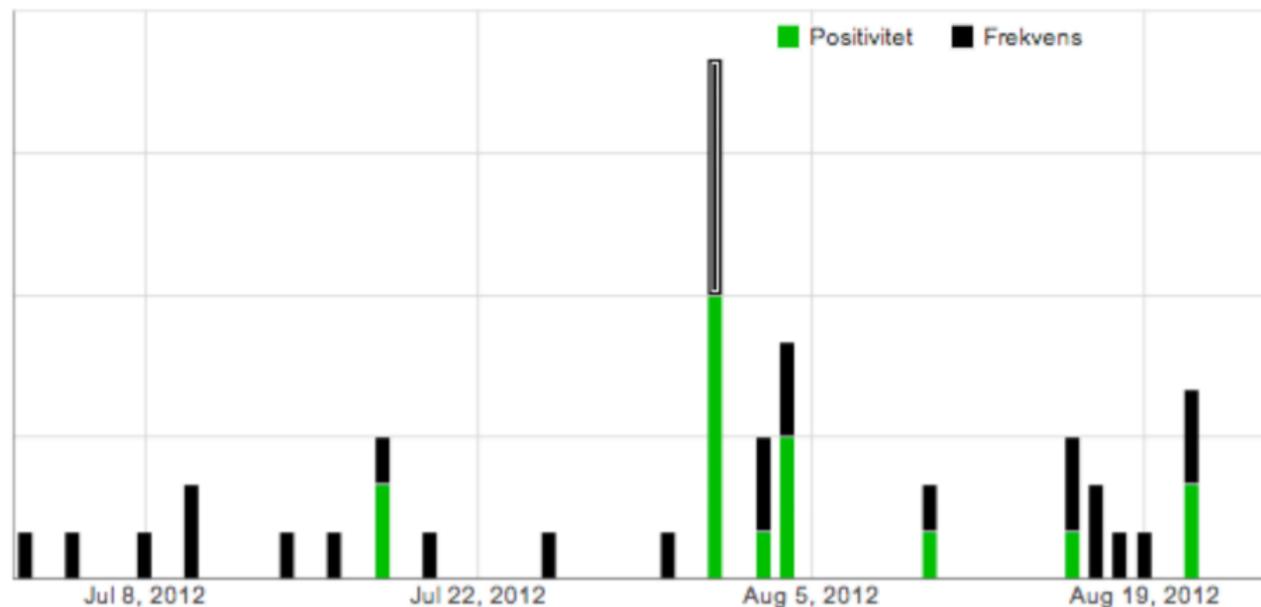
evaluating marketing campaigns

evaluating marketing campaigns

1. marketing campaign

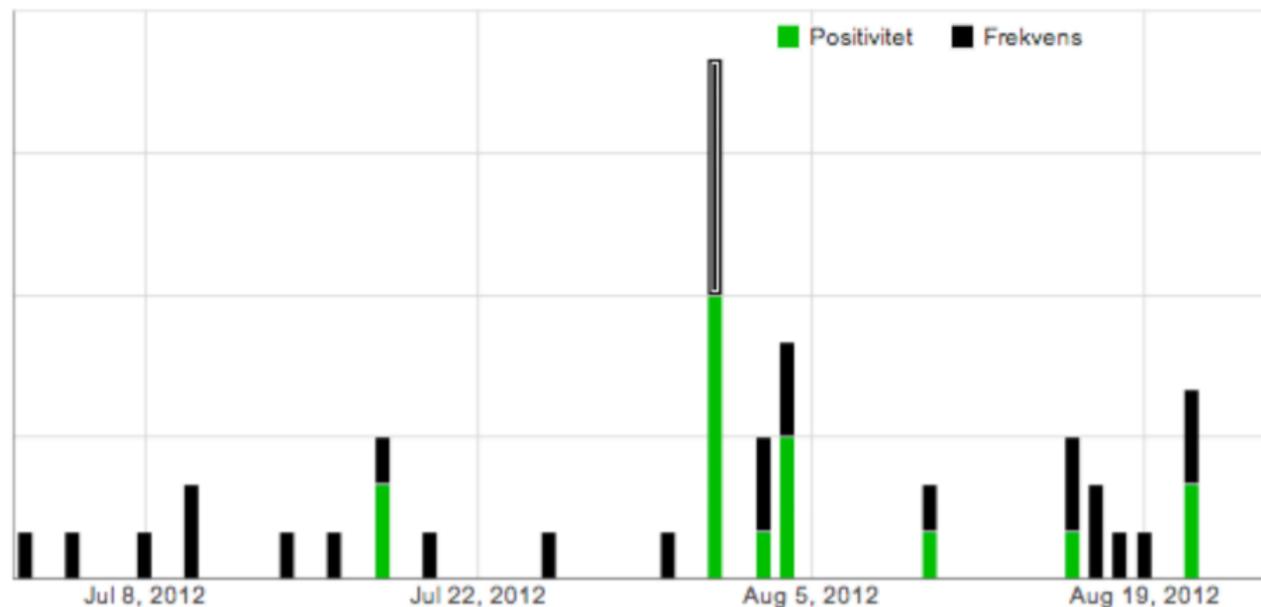# evaluating marketing campaigns

1. marketing campaign
2. freebie with a magazine

evaluating marketing campaigns

1. marketing campaign
2. freebie with a magazine
3. goodie bag

complexity: field study in industrial lab

- complexity because need is complex

complexity: field study in industrial lab

- complexity because need is complex
- complexity because formulation of need is complex

complexity: field study in industrial lab

- complexity because need is complex
- complexity because formulation of need is complex
- complexity because the data to process for fulfilling need is complex

complexity: field study in industrial lab

- complexity because need is complex
- complexity because formulation of need is complex
- complexity because the data to process for fulfilling need is complex
- complexity because packaging what would fulfil need is complex

complexity: field study in industrial lab

- complexity because need is complex
- complexity because formulation of need is complex
- complexity because the data to process for fulfilling need is complex
- complexity because packaging what would fulfil need is complex
- complexity because establishing if need is fulfilled is complex

complexity: field study in industrial lab

- complexity because need is complex
- complexity because formulation of need is complex
- complexity because the data to process for fulfilling need is complex
- complexity because packaging what would fulfil need is complex
- complexity because establishing if need is fulfilled is complex
- complexity because a fulfilled need is not operational or actionable

complexity because need is complex

- information need is almost never complex in mind of the needy customer

complexity because need is complex

- information need is almost never complex in mind of the needy customer
- (or the complexity the customer worries about is often not what is complex)

complexity because need is complex

- information need is almost never complex in mind of the needy customer
- (or the complexity the customer worries about is often not what is complex)
- our job: tell the customer another iteration over need is necessary

supermarkets and csr

supermarkets and csr

health products and kcal labeling

supermarkets and csr

health products and kcal labeling

vs

kitchen technology and cooking trends

supermarkets and csr

simple case: no signal $\longrightarrow$ bad

health products and kcal labeling

vs

kitchen technology and cooking trends

supermarkets and csr

simple case: no signal $\longrightarrow$ bad

health products and kcal labeling

simple case: no signal $\longrightarrow$ good

vs

kitchen technology and cooking trends

supermarkets and csr

simple case: no signal $\longrightarrow$ bad

health products and kcal labeling

simple case: no signal $\longrightarrow$ good

vs

kitchen technology and cooking trends

complex case: plenty of signal $\longrightarrow$ complex analysis

supermarkets and csr

simple case: no signal $\longrightarrow$ bad

health products and kcal labeling

simple case: no signal $\longrightarrow$ good

vs

kitchen technology and cooking trends

complex case: plenty of signal $\longrightarrow$ complex analysis

more complexity is good --- we encourage our
customers to think about complexity in purchase path
rather than follow literal mentions of their brand

complexity because the data to process for fulfilling need is complex

complexity because the data to process for fulfilling need is complex

big data: items not interesting but their occurrence pattern is

complexity because packaging what would fulfil need is complex

visualisation!

complexity because a fulfilled need is not operational or actionable

does anyone in the organisation act on information delivered to them?

complexity because a fulfilled need is not operational or actionable

does anyone in the organisation act on information delivered to them?

more information sources means more complex business, not less work

complexity because establishing if need is fulfilled is complex

depends crucially on previous point --- if no one on receiving end acts on delivery, no validation of effectiveness can be done

complexity because establishing if need is fulfilled is complex

depends crucially on previous point --- if no one on receiving end acts on delivery, no validation of effectiveness can be done

provide that which makes user confident they have what they need

(cf. fizzy drink example)

take home

human linguistic behaviour is wide-ranging and
variable

take home

human linguistic behaviour is wide-ranging and
variable
most everything important is in text somewhere

take home

human linguistic behaviour is wide-ranging and
variable
most everything important is in text somewhere
finding the important stuff is a challenge (a fun
challenge)

take home

human linguistic behaviour is wide-ranging and
variable
most everything important is in text somewhere
finding the important stuff is a challenge (a fun
challenge)
bringing it to the attention of the right people is a
challenge (fun ... maybe)