Think outside the search box

Dr. Tony Russell-Rose FBCS CITP CEng

What's this?

1. randomized controlled trial.pt.
2. controlled clinical trial.pt.
3. randomized.ab.
4. placebo.ab.
5. clinical trials as topic.sh.
6. randomly.ab.
7. trial.ti.
8. 1 or 2 or 3 or 4 or 5 or 6 or 7
9. (animals not (humans and animals)).sh.
10. 8 not 9
11. exp Child/
12. ADOLESCENT/
13. exp infant/
14. child hospitalized/
15. adolescent hospitalized/
16. (child$ or infant$ or toddler$ or adolescen$ or teenage$).tw.

17. or/11-16
18. Child Nutrition Sciences/
19. exp Dietary Proteins/
20. Dietary Supplements/
21. Dietetics/
22. or/18-21
23. exp Infant, Newborn/
24. exp Overweight/
25. exp Eating Disorders/
26. Athletes/
27. exp Sports/
28. exp Pregnancy/
29. exp Viruses/
30. (newborn$ or obes$ or "eating disorder$" or pregnan$ or childbirth or virus$ or influenza).tw.
31. or/23-30
32. 10 and 17 and 22
33. 32 not 31

The secret life of recruiters…

And others

```
1 A01N0025-004/CPC
2 RODENT OR RAT OR RATS OR MOUSE OR
MICE
3 BAIT OR POISON
4 2 AND 3
5 1 OR 4
6 AVERSIVE OR ADVERSIVE OR DETER? OR
REPEL?
7 NONTARGET OR (NON WITH TARGET) OR
HUMAN OR DOMESTIC OR PET OR DOG OR
CAT
8 6 AND 7
9 8 AND 5
10  BITREX OR DENATONIUM OR BITREXENE
OR BITTERANT OR BITTER
11  10 AND 5
12  9 OR 11
```

```
(
    (
        (~Magnolia OR ~Magnolias) AND
        ('~Paul ~Anderson' OR '~Paul ~Thomas' OR paul* OR thom*
OR anders*)
    ) OR
    'jeremy blackman*' OR jeremy OR Jerem* OR
    (
        (~incredible OR ~coincidences) AND
        (role* OR chance* OR life* OR investigat* OR disturb*)
    ) OR
    (
        (~C NOT
            ('~John C' OR
                (
                    (
                        ('~Mister C' AND ~Shamen) OR
                        ('~Tom C' AND ~Scientology)
                    )
                )
            )
        )
    ) OR ~XCs OR ~CV OR ~CVs OR
    (
        (
            (~CKR OR ~CKRs) NOT ('CKR 123' OR lawyer)
        )
    )
)
```

And more…

(topic:"d5f485b9\-aab0\-4d95\-a83b\-28b2decd900c" AND (topic:"ba942587\-854e\-407d\-81dc\-812ee29624bd" OR topic:"805a2281\-8f04\-42b7\-a451\-e2b7daac4db1" OR topic:"1fc2a4f5\-3f23\-4c80\-943c\-42f789d11644" OR topic:"91e4df78\-db55\-4d63\-be65\-2fad1afa77a7" OR organization:"Argos" OR organization:"Sainsbury" OR organization:"Marks and Spencer" OR organization:"Morrisons" OR organization:"Signet" OR organization:"Rapha" OR organization:"Vodafone" OR organization:"Evans Cycle" OR organization:"The Body Shop" OR organization:"Homebase" OR organization:"Topps Tiles" OR organization:"Carphone Warehouse" OR organization:"B\&Q" OR organization:"Tesco" OR organization:"Lidl" OR organization:"Aldi" OR organization:"Waitrose" OR organization:"John Lewis" OR organization:"Dixons" OR organization:"Alliance Boots" OR organization:"AS Watson" OR organization:"Halfords" OR organization:"Vision Express" OR organization:"Pets at Home" OR organization:"Asda" OR organization:"Pret a Manger" OR organization:"Greggs" OR organization:"Majestic Wine" OR organization:"Ocado" OR organization:"Selfridges" OR organization:"Nike" OR organization:"Adidas" OR organization:"Walmart" OR organization:"Woolworths" OR organization:"Debenhams" OR organization:"Halfords" OR organization:"Punch Taverns" OR organization:"Monarch Group" OR organization:"Whitbread" OR organization:"Carnival Cruise" OR organization:"KFC" OR organization:"Yum\!" OR organization:"Costa Coffee" OR organization:"Pizza Hut" OR organization:"Greene King" OR organization:"Carluccio's" OR organization:"Gondola" OR organization:"Starbucks" OR organization:"Sodexo" OR organization:"SSP" OR organization:"Nando's" OR organization:"Avis" OR organization:"Paddy Power" OR organization:"Easyjet" OR organization:"Emaar" OR organization:"Nuffield Health" OR organization:"Safestore" OR organization:"TUI" OR organization:"Merlin Entertainment" OR organization:"Virgin Atlantic" OR organization:"IHG" OR organization:"Virgin Active" OR organization:"Thomas Cook" OR organization:"Odeon Hilton" OR organization:"Pizza Express" OR organization:"Zizzi's" OR organization:"Gemfields" OR organization:"Orlebar Brown" OR organization:"Rebecca Trowell" OR organization:"Spring Studios" OR organization:"Loewe" OR organization:"Jimmy Choo" OR organization:"Burberry" OR organization:"Conde Nast" OR organization:"Zadig \& Voltaire" OR organization:"Prada" OR organization:"Fendi" OR organization:"French Connection" OR organization:"Lyle \& Scott" OR organization:"LK Bennett" OR organization:"Alexander Wang" OR organization:"Sandro" OR organization:"Isabel Marant" OR organization:"Rag \& Bone" OR organization:"J Crew" OR organization:"Maje" OR organization:"Ralph Lauren" OR organization:"Tommy Hilfiger" OR organization:"Michael Kors" OR organization:"Kenzo" OR organization:"Pringle of Scotland" OR organization:"Heidi Klein" OR organization:"Louis Vuitton" OR organization:"Lanvin" OR organization:"Loewe" OR organization:"Chanel" OR organization:"Valentino" OR organization:"Celine" OR organization:"Dior" OR organization:"Alexander McQueen" OR organization:"Victoria Beckham" OR organization:"Sonia Rykel" OR organization:"Bally" OR organization:"Notonthehighstreet" OR organization:"Rupert Sanderson" OR organization:"Dunhill" OR organization:"De Beers" OR organization:"Swarovski" OR organization:"Hermes" OR organization:"Cartier" OR organization:"David Yurman" OR organization:"Mulberry" OR organization:"Tiffany \& Co" OR organization:"Boucheron" OR organization:"Lancel" OR organization:"Marc Jacobs" OR organization:"Longchamp" OR organization:"Kate Spade" OR organization:"Salvatore Ferragamo" OR organization:"Asos" OR organization:"Net a porter" OR organization:"Astley Clarke" OR organization:"Fenwick" OR organization:"Yoox" OR organization:"Harrods" OR organization:"Selfridges" OR organization:"MyTheresa" OR organization:"Lane Crawford" OR organization:"Liberty of London" OR organization:"Gucci" OR organization:"Armarni" OR organization:"Zara" OR organization:"Topshop" OR organization:"Arcadia" OR organization:"Balenciaga" OR organization:"Stella McCartney" OR organization:"Brioni" OR organization:"Bottega Veneta" OR organization:"Christopher Kane" OR organization:"Anya Hindmarch" OR organization:"Business of Fashion" OR organization:"3i" OR organization:"Active Private Equity" OR organization:"Advent International" OR organization:"Advent" OR organization:"Apollo Global Management" OR organization:"Apollo" OR organization:"Apax" OR organization:"Apax" OR organization:"Augmentum Capital" OR organization:"Bain" OR organization:"Endless" OR organization:"BC Partners" OR organization:"Better Capital" OR organization:"HuttonCollins" OR organization:"Hutton Collins" OR organization:"Bridgepoint" OR organization:"Business Growth Fund" OR organization:"CDR" OR organization:"Darwin" OR organization:"Duke Street Capital" OR organization:"Electra Partners" OR organization:"Exponent" OR organization:"European Capital" OR organization:"General Atlantic" OR organization:"Hilco" OR organization:"Inflexion" OR organization:"KKR" OR organization:"LGV" OR organization:"Lion" OR organization:"Magenta Partners" OR organization:"Montagu" OR organization:"Octopus" OR organization:"Octopus Capital" OR organization:"Piper Private Equity" OR organization:"Piper" OR organization:"Sun Capital" OR organization:"TA Associates" OR organization:"Warburg Pincus" OR organization:"Charterhouse" OR organization:"Primary Capital" OR organization:"Bowmark" OR organization:"Beringea" OR organization:"Business Growth Fund" OR organization:"Bridgepoint" OR organization:"Caird Capital" OR organization:"CapVest" OR organization:"Cerberus Capital" OR organization:"CBPE" OR organization:"CCMP" OR organization:"CDR" OR organization:"Change Capital" OR organization:"Cinven" OR organization:"CVC" OR organization:"Duke Street" OR organization:"EC1" OR organization:"Electra" OR organization:"Encore" OR organization:"EQT" OR organization:"Equistone" OR organization:"Graphite Capital" OR organization:"Caledonian" OR organization:"Kings Park Capital" OR organization:"L Capital" OR organization:"LDC"

Agenda

- Understanding the problem
  - User needs, goals & tasks
- Design exploration
  - Alternative solutions
- Evaluation
  - Moving the needle
- Exploitation
 - Gaining traction
- Summary

Professional search

- Keyword search is not good enough
  - 'Advanced' search is anything but
- Patent, healthcare, recruitment, media...
  - Professional search needs to be:
    - Comprehensive
   - Repeatable
    - Transparent

Healthcare Information

- Evidence based medicine
- *"A systematic review summarises the results of healthcare studies (controlled trials) and provides a high level of evidence on the effectiveness of healthcare interventions. Judgments may be made about the evidence and inform recommendations for healthcare."*
- *Cochrane Consumer Network,* https://consumers.cochrane.org/what-systematic-review

Big data in healthcare
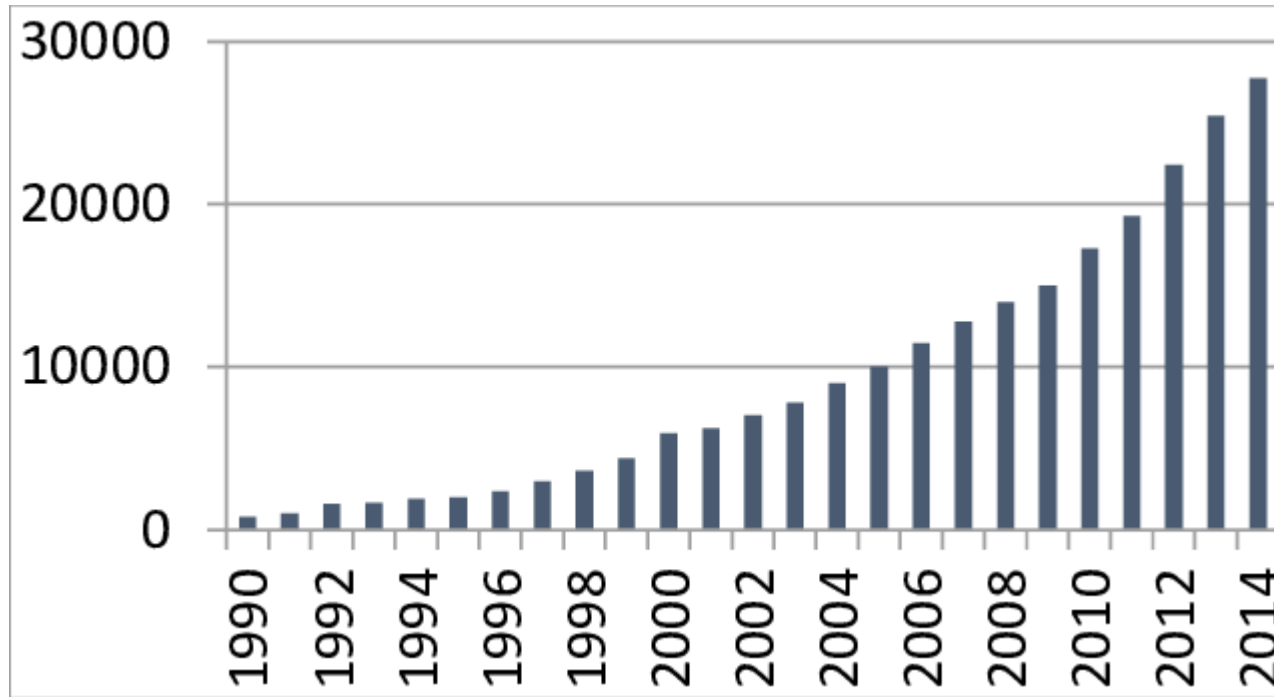
Bekhuis T, Tseytlin E, Mitchell KJ. A Prototype for a Hybrid System to Support Systematic Review Teams: A Case Study of Organ Transplantation. *Proceedings IEEE International Conference on Bioinformatics and Biomedicine*. 2015;2015:940-947.

Big data in healthcare
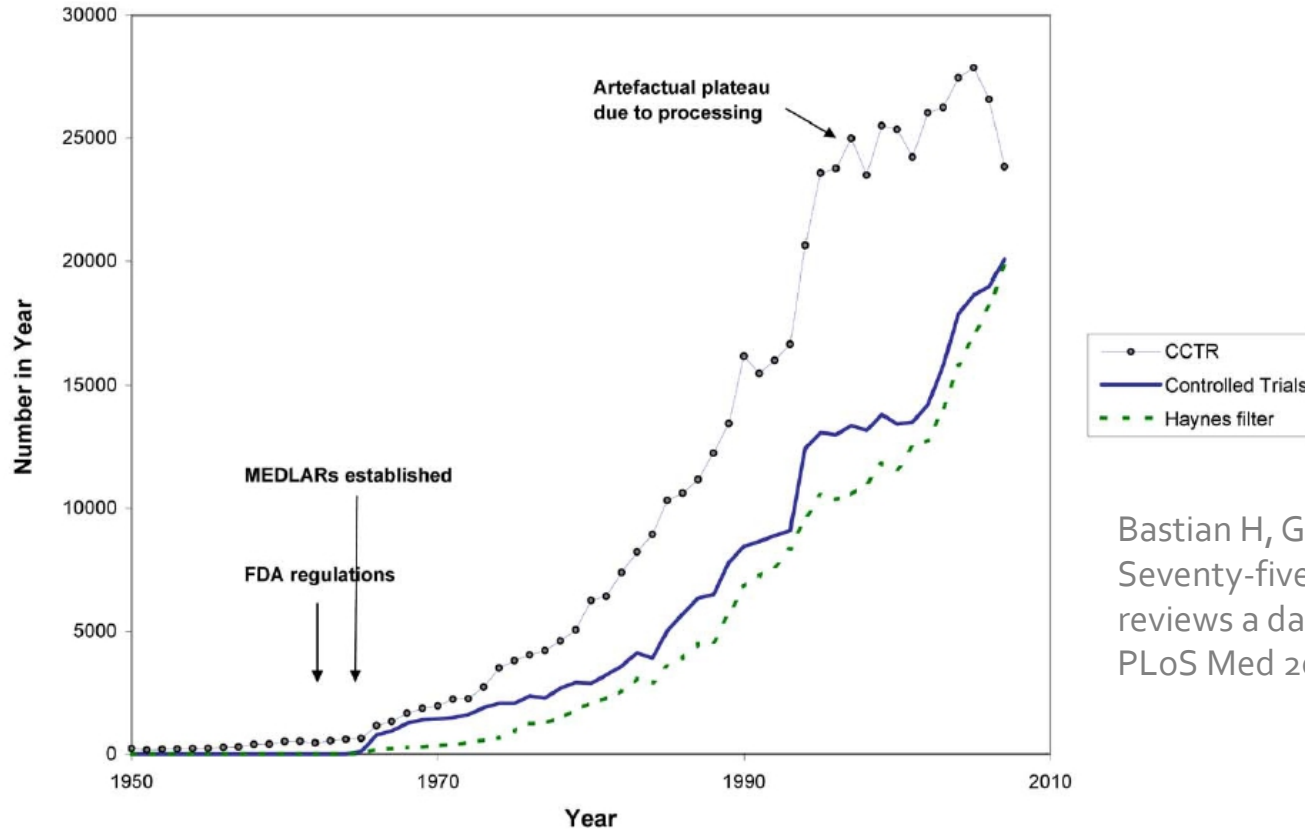
Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med 2010 Sep 21;7(9)

An imperfect process

- *"Errors in the search strategy of a systematic review may undermine the integrity of the evidence base used in the review"*
    - Sampson M, McGowan J. Errors in search strategies were identified by type and frequency. J Clin Epidemiol 2006 Oct;59(10):1057-1063

- 63 MEDLINE search strategies examined
    - 90.5% contained ≥1 errors
    - 82.5% contained errors that could potentially lower recall of relevant studies
    - The most common search errors were
        - missed MeSH terms (44.4%)
        - unwarranted explosion of MeSH terms (38.1%)
        - irrelevant MeSH or free text terms (28.6%)
        - Missed spelling variants, combining MeSH and free text terms in the same line, and failure to tailor the search strategy for other databases (20.6%)
        - Logical operator error occurred in 19.0% of searches

Quiz time! 2008, or 2018?

Think o

Quiz time! 2008, or 2018?

Learning from other disciplines

- Search strategies are:
  - Complex, logical constructs
  - Read many more times that they are written
  - Exhibit abstraction, modularity, encapsulation
  - Maintained, shared, optimised...
- *"For the IBM PC in the early 1980s, the software development process was text edit, compile, write down the errors, and debug with your eyes."*
  - The History of Visual Development Environments, *https://www.mendix.com/blog/the-history-of-visual-development-environments-imagine-theres-no-ides-its-difficult-if-you-try/*

Thinking outside the search box

- What if we could *visualize* search strategies?
    - Use metaphors the user already understands
    - Separate keyword selection from query manipulation
    - Provide a 'scratchpad' & 'debugging aids'
    - Provide instant feedback
    - Support abstraction (levels of composition)
    - Provide query suggestions

Anick et al (1989), Fishkin & Stone (1995)
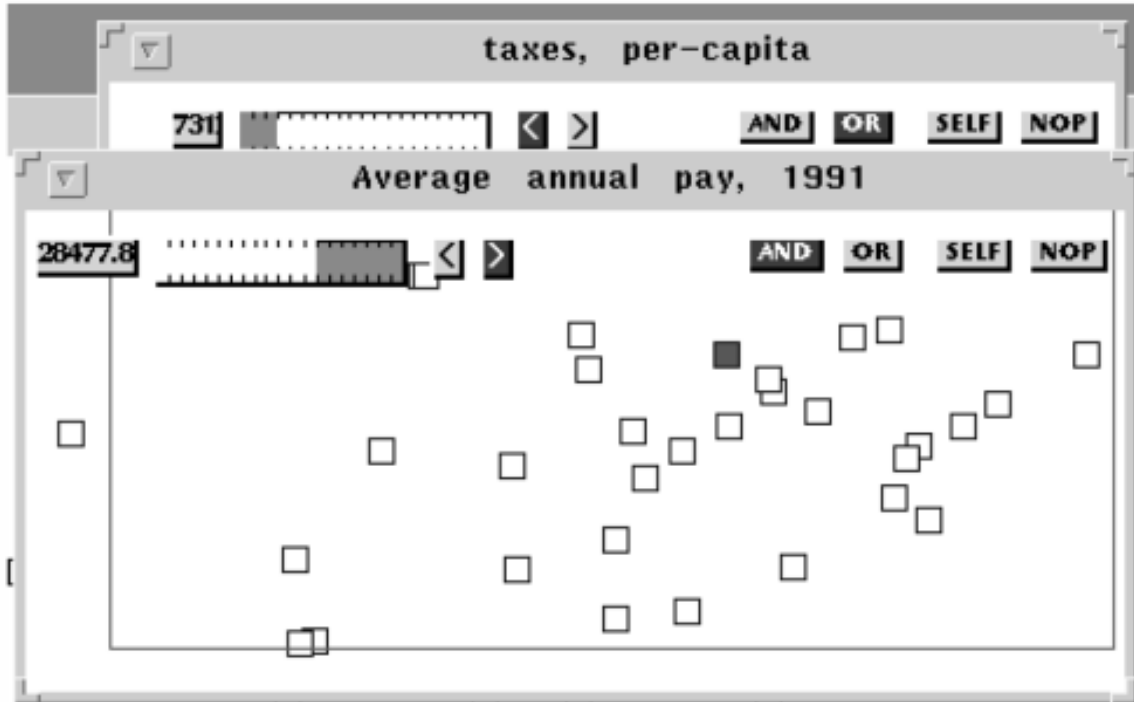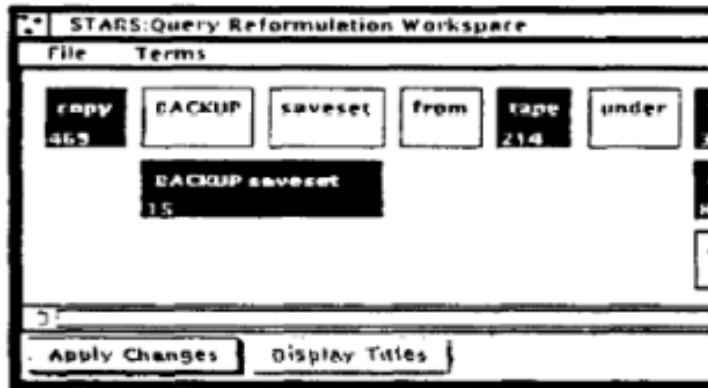
Jones (1998), Yi et al (2005)
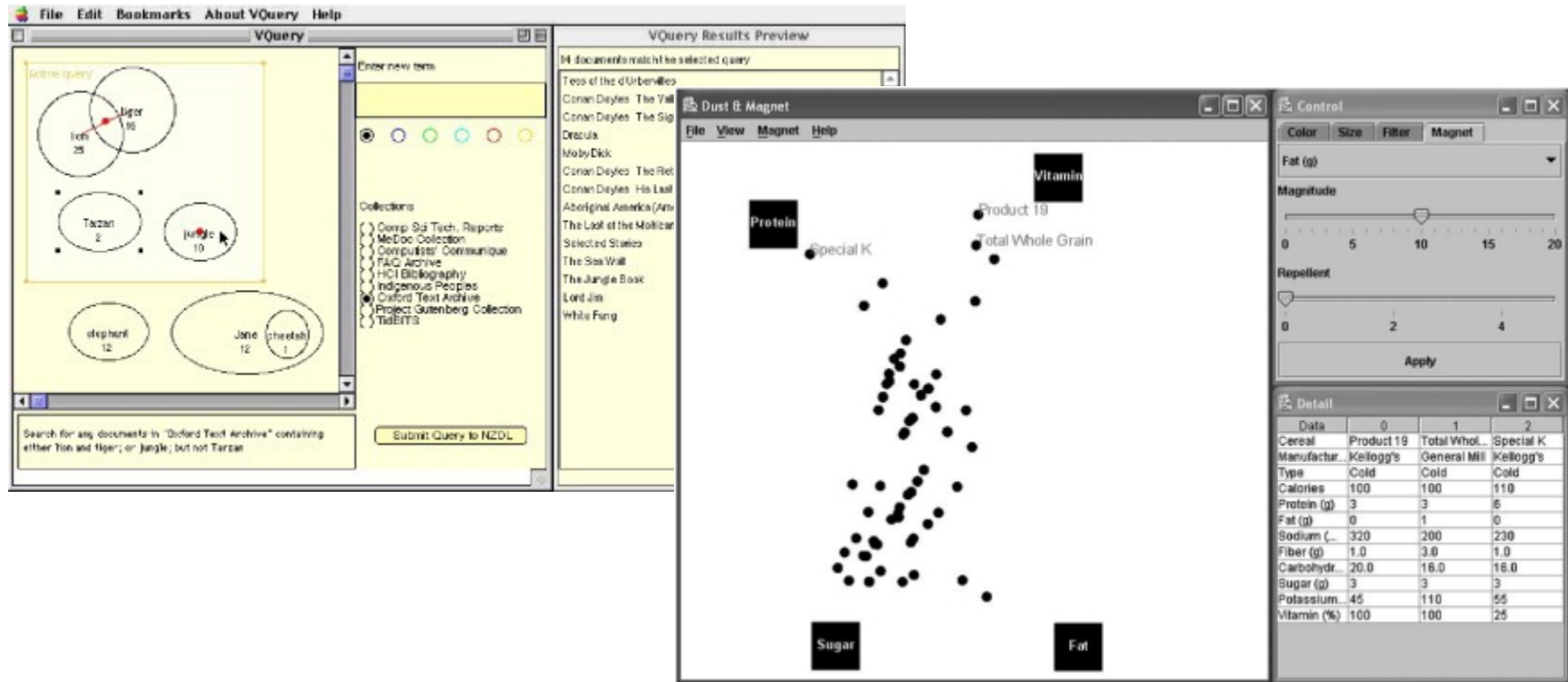
Nitsche & Nurnberger (2013), Boolify (2014?)
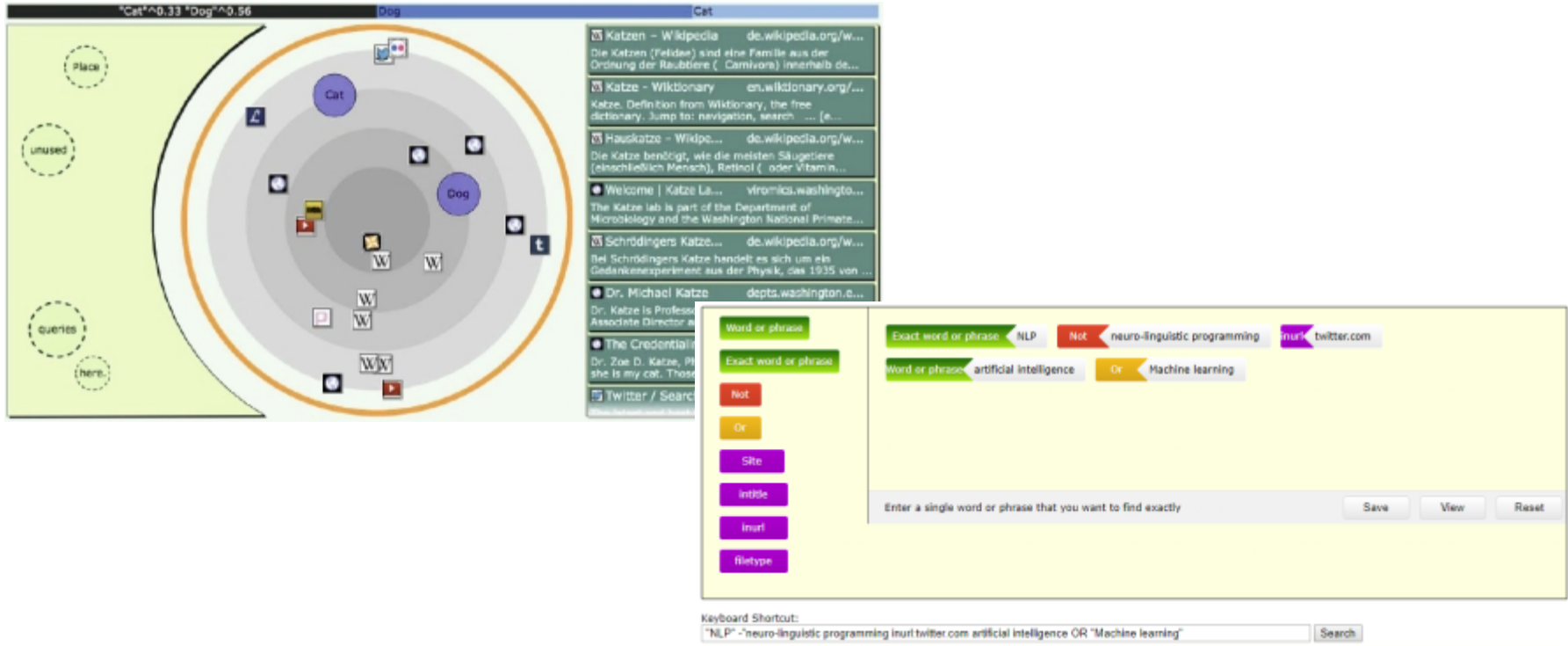
Query formulation

- Concepts are objects on a 2d canvas
- Relationships expressed via direct manipulation
  - Only syntactically correct expressions
  - Make the semantics transparent
  - Support arbitrary levels of abstraction
- Support platform-specific operators
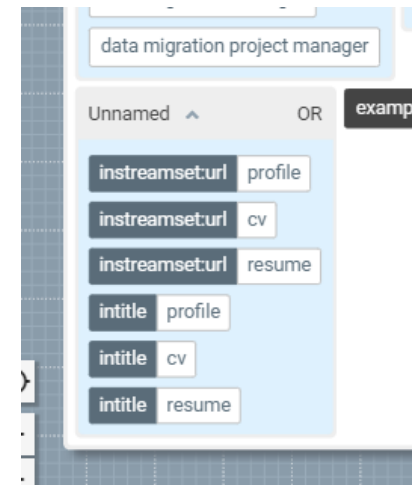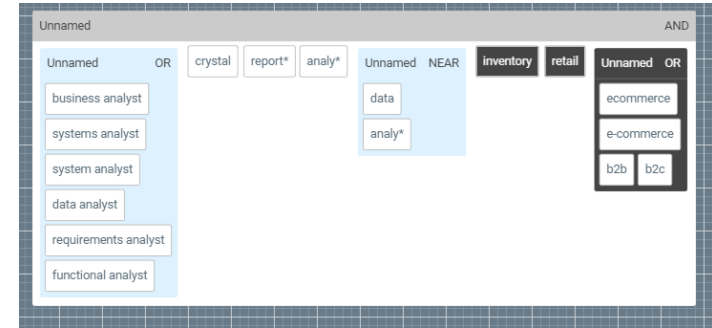
Query refinement

- Cut, copy, paste
- Single or multiple select
  - Lasso
- Undo / redo
- Mapping metaphor
  - Conceptual zoom
  - Overview+detail
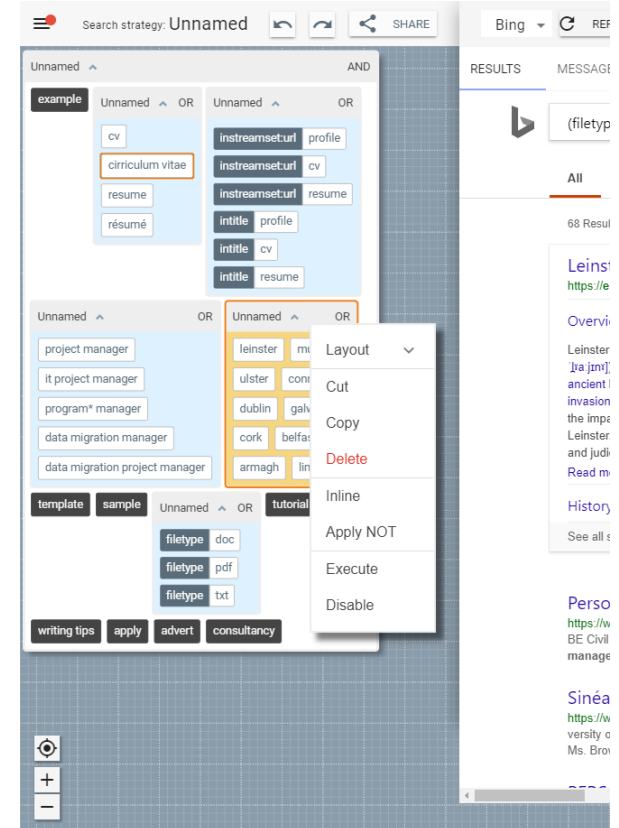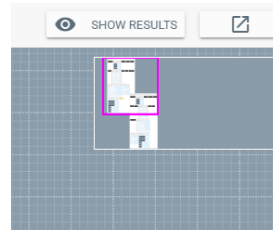
Executing queries

- Semantic layer:
  - Google, Bing, PubMed…
  - …your DB?
- Universal representation?
  - Mappings may be ambiguous
- Boolean string I/O
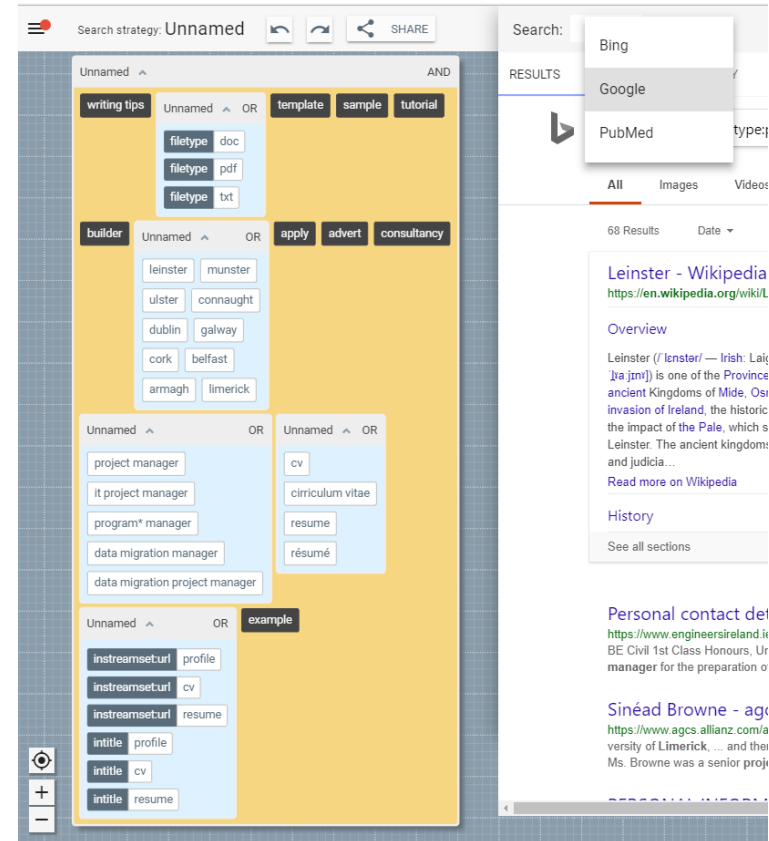
But what about content?

- Term selection is common source of inefficiency & error

  - Apply query suggestions?
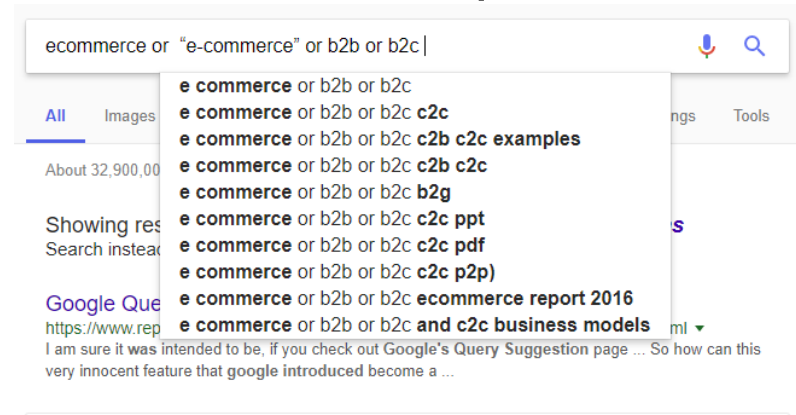
  - *("business analyst" or "systems analyst" or "system analyst" or "data analyst" or "requirements analyst" or "functional analyst") and crystal and report* and analy* and data near analy* and not inventory and not retail and not (ecommerce or "e-commerce" or b2b or b2c)*

Suggestions in context

Query suggestions

- No access to source documents(!)
- Various methods:
  - Global, ontology lookup
    - DBPEDIA, WEBISA, MeSH, etc.
  - Global, word embeddings
    - Wikipedia, GoogleNews, PubMed, bespoke
  - Local, document clustering
    - Carrot2 + PubMed search result snippets

NLP architecture
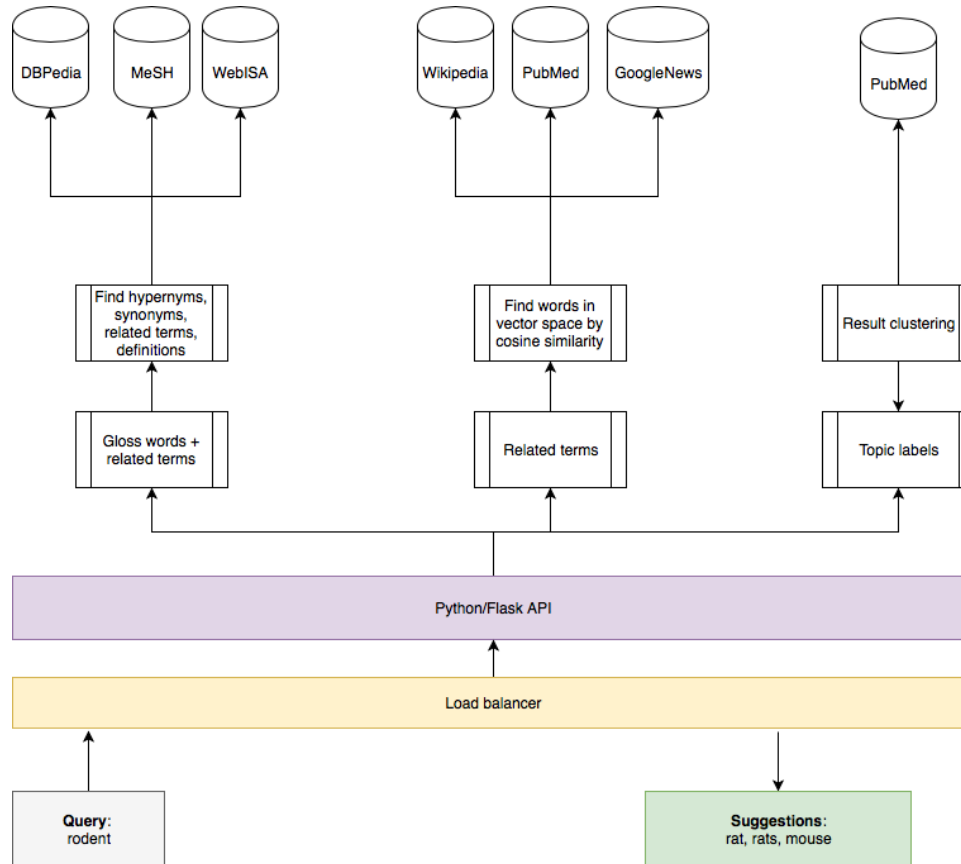
T.G. Russell-Rose and Phil Gooch ,"2dSearch: a visual approach to search strategy formulation", Proceedings of DESIRES: Design of Experimental Search & Information REtrieval Systems, Bertinoro, Italy, 28-31 August 2018

Evaluation

- Query suggestion != query expansion
  - Relevance ranking less important
  - Terms are added to a specific (sub) expression
  - Need to consider individual candidate terms
- Methodology: can we (re)generate published strategies?
  - CLEF 2017 eHealth Lab

Evaluation (ontology terms)

Service	P	R	F
DBPEDIA	0.017	**0.040**	**0.024**
WEBISA	0.001	0.003	0.002
MeSH	**0.045**	0.012	0.019
BNF	0.002	0.001	0.001

Beel, J., & Langer, S. (2015, September). A comparison of offline evaluations, online evaluations, and user studies in the context of research-paper recommender systems. In International Conference on Theory and Practice of Digital Libraries (pp. 153-168). Springer, Cham.

Evaluation (word embeddings)

Model	P	R	F
Word2vec (News)	0.016	0.025	0.019
GloVe (Wikipedia)	0.019	0.030	0.024
Word2vec (PubMed)	0.028	0.042	0.034
FastText (Wikipedia)	0.024	0.038	0.029
Word2vec (PubMed unigram)	0.031	0.047	0.037
Word2vec (PubMed trigram)	**0.035**	**0.052**	**0.042**

T.G. Russell-Rose and Phil Gooch ,"2dSearch: a visual approach to search strategy formulation", Proceedings of DESIRES: Design of Experimental Search & Information REtrieval Systems, Bertinoro, Italy, 28-31 August 2018

But what about content (reuse)?

- Open content repository:
  - Templates and best practices
  - Library of shared strategies (aka 'GitHub')
  - Community of practice (aka 'SO')

Technology: Search strategy formulation

Content: Samples & templates

Commu nity: Sharing & review
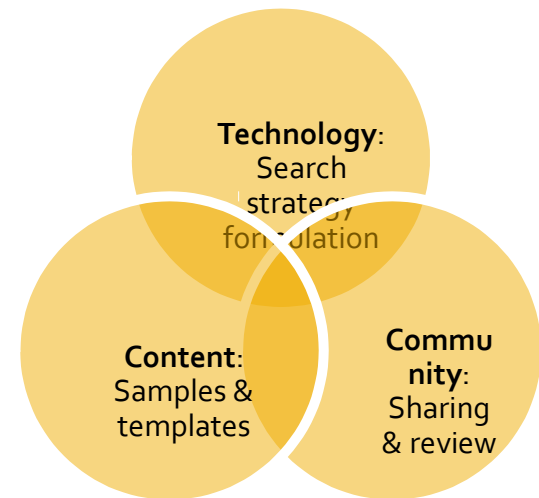
But the hardest part…

- If you build it, they won't come
  - "*Most startups don't fail at building a product. Most startups fail at getting traction*"
    - Gabriel Weinberg

A taxonomy of traction …

Traction
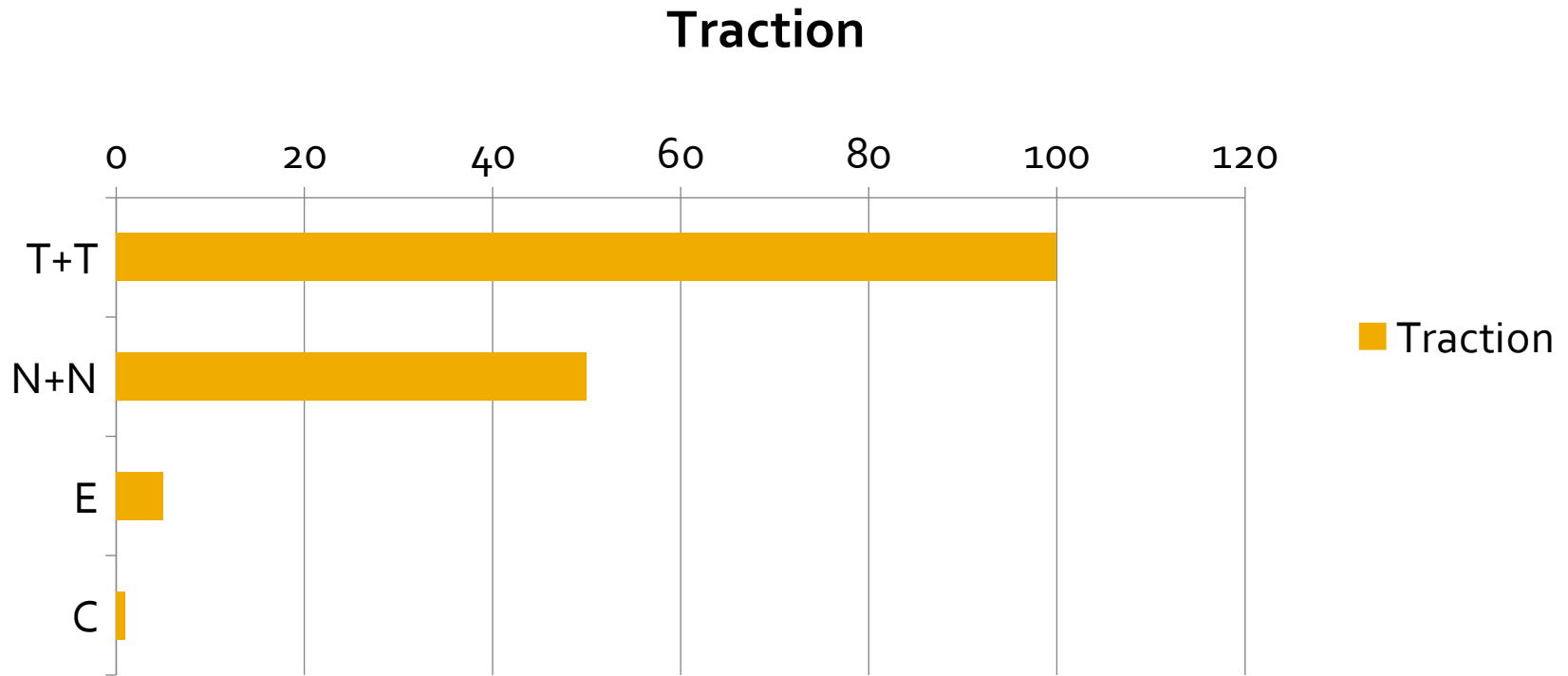
But the hardest part…

- What is 'it' anyway?
  - Pick a problem small enough to solve (and ignore the rest)
- Having competitors is GOOD
  - Proof of market
  - Promotes awareness of the problem/solution
  - Helps you identify what else you need to complete the product
  - Gives a realistic pricing model

Further work

- Formal user-centric evaluation
  - App in context (e.g. vs traditional query builder)
  - Quality of query suggestions
- R&D projects:
  - (semi) automated strategy translation
  - (semi) automated strategy generation (from job specs)
  - Combining set retrieval with ranked retrieval
  - De-duping, optimising, comparing…

Summary

- Reimagining 'advanced search'
  - Eliminate syntactic error
  - Make semantics transparent
  - Provide support for reuse, optimisation, best practice
- Looking for:
  - Early adopters, R&D collaborations
  - Evaluation partners
- Try it out: https://www.2dsearch.com

Thank you

Dr. Tony Russell-Rose, FBCS CITP CEng
Founder, 2dSearch
Director, UXLabs
RAE Visiting Professor, Essex University

- Web: http://www.2dsearch.com
- Email: info@2dsearch.com
- Blog: http://isquared.wordpress.com
- LinkedIn: http://uk.linkedin.com/in/tonyrussellrose
- Twitter: @tonygrr, @2d_search