Think outside the search box

Dr. Tony Russell-Rose FBCS CITP CEng
Professional search

- We all need search but...
  - Some need it more than others
  - Patent, healthcare, recruitment, media...
    - Keyword search is not good enough
  - Professional search needs to be:
    - Comprehensive
    - Repeatable
    - Transparent

Think outside the search box – www.2dsearch.com
Query builders
A healthcare search strategy

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 adolescent$ 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

Recruitment search strategies

Java AND (Design OR develop OR code OR Program) AND ("* Engineer" OR MTS OR "* Develop*" OR Scientist OR technologist) AND (J2EE OR Struts OR Spring) AND (Algorithm OR "Data Structure" OR PS OR "Problem Solving")

Think outside the search box – www.2dsearch.com
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)
  
  )
  
  }
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

Growth in systematic reviews over 25 years (N=233,785 MEDLINE citations)

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)
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) is an evidence-based minimum set of items aimed at helping authors to report a wide array of systematic reviews and meta-analyses that assess the benefits and harms of a health care intervention.

https://en.wikipedia.org/wiki/Preferred_Reporting_Items_for_Systematic_Reviews_and_Meta-Analyses

What’s wrong with Boolean strings?

- Ineffective for communicating *structure*
  - Visual cues needed
- Poor scalability
  - No abstraction
- Error-prone
  - Syntactic + semantic errors


Think outside the search box – www.2dsearch.com
Auditing published search strategies

- 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
  - 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

Think outside the search box – www.2dsearch.com
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.”
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
Nitsche & Nurnberger (2013), Boolify (2014?)
 BOOLIO, Search Whiteboard, Searchrefiner

Think outside the search box – www.2dsearch.com
Concepts are objects on a 2D canvas
- Relationships via direct manipulation
  - Syntactically correct expressions
  - Transparent semantics
  - Arbitrary levels of abstraction
- ‘Lego’ analogy
  - Optimisation thru exploration
    - Execute, enable/disable
    - Hit counts
Executing queries

- Integrations
  - 8 databases & counting...
- Universal representation
  - Auto-translation
  - Error-checking
- Interoperability with legacy formats
  - Boolean string I/O

Think outside the search box – www.2dsearch.com
That healthcare strategy...
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)

Think outside the search box – www.2dsearch.com
Suggestions in context

Think outside the search box – www.2dsearch.com
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
Sharing, reuse, best practice

- Open content repository:
  - Templates and best practices
  - Library of shared strategies (aka ‘GitHub’)
  - Community of practice (aka ‘Stack Overflow’)

Think outside the search box – www.2dsearch.com
Further work

- **User experience**
  - Visual methods vs. traditional query builders
  - Quality of query suggestions

- **Data science**
  - Semi-automated strategy generation
  - Investigation of ConceptNet, Wikidata, etc.
  - Combining set retrieval with ranked retrieval
  - Contextual language modelling (BERT, ELMo, etc.)
Summary

- Reimagining ‘advanced search’
  - Eliminate syntactic error
  - Make semantics transparent
  - Provide support for reuse, optimisation, best practice

- Try it for yourself: [https://www.2dsearch.com](https://www.2dsearch.com)
  - Tutorial videos: [https://www.2dsearch.com/help](https://www.2dsearch.com/help)
Dr. Tony Russell-Rose, FBCS CITP CEng
Senior Lecturer, Goldsmiths
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