Searching fast & slow

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
A01N0025-004/CPC

RODENT OR RAT OR RATS OR MOUSE OR MICE

BAIT OR POISON

2 AND 3

1 OR 4

AVERSIVE OR ADVERSIVE OR DETER? OR REPEL?

NONTARGET OR (NON WITH TARGET) OR HUMAN OR DOMESTIC OR PET OR DOG OR CAT

6 AND 7

8 AND 5

BITREX OR DENATONIUM OR BITREXENE OR BITTERANT OR BITTER

10 AND 5

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)

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
Peer-reviewing search strategies (McGowan et al. 2016, J Clin Epidemiology)
Search strategy development (Russell-Rose & Shokraneh, Journal of Library UX)

- **Word Processors**
  - Auto-conversion of straight to smart quotations
  - Auto-correction of truncation or wildcards: Creates duplicates
  - Auto-space-removal of shared documents
  - Need external program such as Microsoft Office or GoogleDocs

- **PDFs**
  - Breaking paragraphs into lines when copies
  - Cannot be edited (without a installing another program)
  - Need external program such as Microsoft Office or GoogleDocs
Professional search practice

- Reproducibility of searches
- Training beginners
- Translating search strategies
  - PolyGlot
- Sharing search strategies (externally)
  - Unavailability of supplementary materials (Evangelou et al. 2005, FACEB J)
  - MedTerm Assist (Pittsburgh University), Search Strategy Library (Gulhane, 2019)
  - ISSG Search Filters, Search Blocks on Google Sites
  - Institutional Repositories or OSF
- Living Search Strategies (Shokraneh & Russell-Rose 2020, J Clin Epidemiol)
  - COVID-19
Quiz time! 2010 or 2020?
Quiz time! 2010 or 2020?
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.”
Searching fast & slow

**Slow thinking**
- Procedural
- Static
- Monolithic
- Strings

**Fast thinking**
- Declarative
- Interactive
-Executable
- Objects
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

Think outside the search box – www.2dsearch.com
Query formulation

- 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

Think outside the search box – www.2dsearch.com
Executing queries

- Integrations
  - 9 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...
That patent strategy...
Design principles

1. Support **transparency** in the **mapping** between logical structure and physical structure
2. Adopt **scalable formalisms** that accommodate complexity
3. Delegate lower-level syntactic operations to system functions
4. Provide **real-time feedback** on query effectiveness
5. Provide support for **collaboration** and team working

Russell-Rose & MacFarlane, EARS 2020: The 3rd International Workshop on ExplainAble Recommendation and Search2020
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
Reimagining ‘advanced search’
- Eliminate syntactic error, make semantics transparent
- Provide support for reuse, optimisation, best practice

Key transformations:
- Procedural → declarative
- Static → interactive
- Monolithic → executable
- Strings → objects

Try it for yourself: https://www.2dsearch.com
- Tutorial videos: https://www.2dsearch.com/help
Think outside the search box – www.2dsearch.com