Rethinking 'Advanced Search': A New Approach to Complex Query Formulation

Chapter · April 2019

DOI: 10.1007/978-3-030-15719-7_31

3 authors:

Tony Russell-Rose
2dSearch
96 PUBLICATIONS   **2,160** CITATIONS

Jon Chamberlain
University of Essex
34 PUBLICATIONS   **218** CITATIONS

Udo Kruschwitz
University of Essex
154 PUBLICATIONS   **864** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   MultiLing 2017 in EACL 2017 View project

Project   Enterprise/Site/Intranet Search View project

Rethinking 'Advanced Search': A New Approach to Complex Query Formulation

Tony Russell-Rose[1], Jon Chamberlain[2], and Udo Kruschwitz[2]

[1] UXlabs Ltd, Brunel House, 340 Firecrest Ct, Centre Park, Warrington WA1 1RG
tgr@uxlabs.co.uk
[2] School of Computer Science and Electronic Engineering, University of Essex,
Wivenhoe Park, Colchester, Essex, UK, CO4 3SQ <jchamb|udo>@essex.ac.uk

Abstract. Knowledge workers such as patent agents, recruiters and media monitoring professionals undertake work tasks where search forms a core part of their duties. In these instances, the search task often involves the formulation of complex queries expressed as Boolean strings. However, creating effective Boolean queries remains an ongoing challenge, often compromised by errors and inefficiencies. In this demo paper, we present a new approach to query formulation in which concepts are expressed on a two-dimensional canvas and relationships are articulated using direct manipulation. This has the potential to eliminate many sources of error, makes the query semantics more transparent, and offers new opportunities for query refinement and optimisation.

Keywords: query formulation · advanced search · Boolean · search visualisation · professional search

1 Introduction

Many knowledge workers rely on the effective use of search applications in the course of their professional duties [6]. Patent agents, for example, depend on accurate prior art search as the foundation of their due diligence process [10]. Similarly, recruitment professionals rely on Boolean search as the basis of the candidate sourcing process [8], and media monitoring professionals routinely manage thousands of Boolean expressions on behalf their client briefs [12].

The traditional solution is to formulate complex Boolean expressions consisting of keywords, operators and search commands, such as that shown in Figure 1. However, the practice of using Boolean strings to articulate complex information needs suffers from a number of fundamental shortcomings [9]. First, it is poor at communicating structure: without some sort of physical such as indentation, parentheses and other delimiters can become lost among other alphanumeric characters. Second, it scales poorly: as queries grow in size, readability becomes progressively degraded. Third, they are error-prone: even if syntax checking is provided, it is still possible to place parentheses incorrectly, changing the semantics of the whole expression.

(cv OR "cirriculum vitae" OR resume OR "resum") (filetype:doc OR file-type:pdf OR filetype:txt) (inurl:profile OR inurl:cv OR inurl:resume OR initile:profile OR intitle:cv OR initile:resume) ("project manager" OR "it project manager" OR "program* manager" OR "data migration manager" OR "data migration project manager") (leinster OR munster OR ulster OR connaught OR dublin) -template -sample -example -tutorial -builder -"writing tips" -apply -advert -consultancy

Fig. 1: An example from the Boolean Search Strings Repository

To mitigate these issues, many professionals rely on previous examples of best practice. Recruitment professionals, for example, draw on repositories such as the Boolean Search Strings Repository[3] and the Boolean String Bank[4]. However, these repositories store content as unstructured text strings, and as such their true value as source of experimentation and learning may never be fully realized.[5]

2dSearch[6] offers an alternative approach. Instead of formulating Boolean strings, queries are expressed by combining objects on a two-dimensional canvas and relationships are articulated using direct manipulation. This eliminates many sources of syntactic error, makes the query semantics more transparent, and offers further opportunities for query refinement and optimisation.

2   Related Work

The application of data visualisation to search query formulation can offer significant benefits, such as fewer zero-hit queries, improved query comprehension and better support for exploration of an unfamiliar database [3]. An early example is that of Anick et al. [1], who developed a two-dimensional graphical representation of a users natural language query that supported reformulation via direct manipulation. Fishkin and Stone [2] investigated the application of direct manipulation techniques to database query formulation, using a system of lenses to refine and filter the data. Jones [4] developed a query interface to the New Zealand Digital Library which uses Venn diagrams and integrated query result previews.

A further example is Yi et al. [13], who applied a dust and magnet metaphor to multivariate data visualization. Nitsche and Nurnberger[5]] developed a system based on a radial user interface that supports phrasing and interactive visual refinement of vague queries. A further example is Boolify[7], which provides a drag and drop interface to Google. More recently, de Vries et al [11] developed a system which utilizes a visual canvas and elementary building blocks to allow users

---

[3] https://booleanstrings.ning.com/forum/topics/boolean-search-strings-repository, accessed 10 Oct 2018.

[4] https://scoperac.com/booleanstringbank, accessed 10 Oct 2018.

[5] http://booleanblackbelt.com/2016/01/the-most-powerful-boolean-search-operator, accessed 10 Oct 2018.

[6] https://2dsearch.com, accessed 24 Oct 2018.

[7] https://www.kidzsearch.com/boolify/, accessed 23 Oct 2018

to graphically configure a search engine. 2dSearch differs from the prior art in offering a database-agnostic approach with automated query suggestions and support for optimising, sharing and re-using query templates and best practices.

3   Design Concept

At the heart of 2dSearch is a graphical editor which allows the user to formulate queries as objects on a two-dimensional canvas. Concepts can be simple keywords or attribute:value pairs representing controlled vocabulary terms or database-specific search operators. Concepts can be combined using Boolean (and other) operators to form higher-level groups and then iteratively nested to create expressions of arbitrary complexity. Groups can be expanded or collapsed on demand to facilitate transparency and readability.
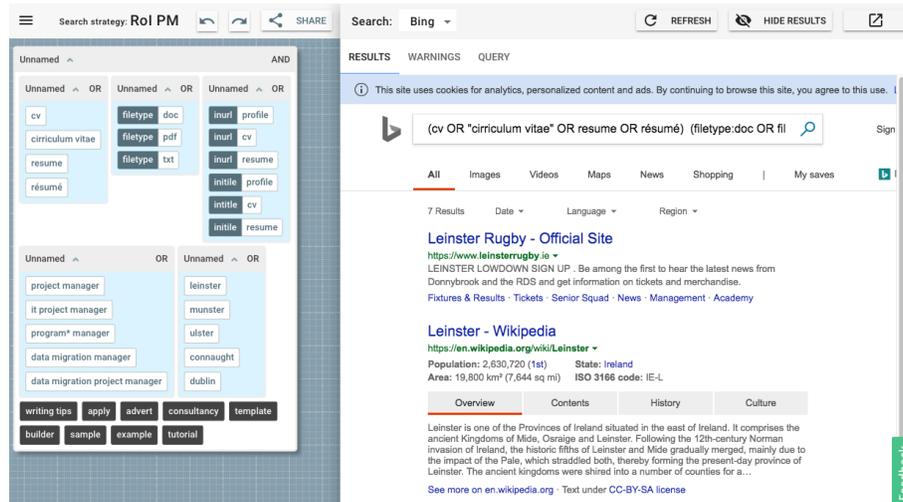

Fig. 2: The 2dSearch app showing query canvas (left) and search results pane (right).

The application consists of two panes (see Figure 2): a query canvas and a search results pane (which can be resized or detached in a separate window). The canvas can be resized or zoomed, and features an 'overview' widget to allow users to navigate to elements that may be outside the current viewport. Adopting design cues from Google's Material Design language[8], a sliding menu is offered on the left, providing file I/O and other options. This is complemented by a navigation bar which provides support for document-level functions such as naming and sharing queries.

Although 2dSearch supports creation of complex queries from a blank canvas, its value is most readily understood by reference to an example such as that

---

[8] https://material.io

of Figure 1, which is intended to find social profiles for data migration project managers located in Dublin. Although relatively simple, this query is still difficult to interpret, optimise or debug. However, when opened with 2dSearch, it becomes apparent that the overall expression consists of a conjunction of OR clauses (nested blocks) with a number of specialist search operators (dark blue) and negated terms (white on black). To edit the expression, the user can move terms using direct manipulation or create new groups by combining terms. They can also cut, copy, delete, and lasso multiple objects. If they want to understand the effect of one group in isolation, they can execute it individually. Conversely, if they want to remove one element from consideration, they can disable it. In each case, the effects of each operation are displayed in real time in the adjacent search results pane.

2dSearch functions as a meta-search engine, so is in principle agnostic of any particular search technology or platform. In practice however, to execute a given query, the semantics of the canvas content must be mapped to the API of the underlying database. This is achieved via an abstraction layer or set of 'adapters' for common search platforms such as Bing, Google, PubMed, Google Scholar, etc. These are user selectable via a drop-down control.

Support for query optimisation is provided via a 'Messages' tab on the results pane. For example, if the user tries to execute via Bing a query string containing operators specific to Google, an alert is shown listing the unknown operators. 2dSearch also identifies redundant structure (e.g. spurious brackets or duplicate elements) and supports comparison of canonical representations. Query suggestions are provided via an NLP services API which utilises various Python libraries (for word embedding, keyword extraction, etc.) and SPARQL endpoints (for linked open data ontology lookup)[7].

4   Summary and Further Work

2dSearch is a framework for search query formulation in which information needs are expressed by manipulating objects on a two-dimensional canvas. Transforming logical structure into physical structure mitigates many of the shortcomings of Boolean strings. This eliminates syntax errors, makes the query semantics more transparent and offers new ways to optimise, save and share best practices. In due course, we hope to engage in a formal, user-centric evaluation, particularly in relation to traditional query builders. We are currently engaging in an outreach programme and invite subject matter experts to work with us in building repositories of curated (or user generated) examples and templates.

Adopting a database-agnostic approach presents challenges, but it also offers the prospect of a *universal* framework in which information needs can be articulated in a *generic* manner and the task of mapping to an underlying database can be delegated to platform-specific adapters. This could have profound implications for the way in which professional search skills are taught, learnt and applied.

References

1. Anick, P.G., Brennan, J.D., Flynn, R.A., Hanssen, D.R., Alvey, B., Robbins, J.M.: A direct manipulation interface for boolean information retrieval via natural language query. In: Proceedings of the 13th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 135–150. SIGIR '90, ACM, New York, NY, USA (1990). https://doi.org/10.1145/96749.98015
2. Fishkin, K., Stone, M.C.: Enhanced dynamic queries via movable filters. pp. 415–420. ACM Press (1995)
3. Goldberg, J.H., Gajendar, U.N.: Graphical condition builder for facilitating database queries. U.S. Patent No. 7,383,513. 3 (2008)
4. Jones, S.: Graphical query specification and dynamic result previews for a digital library. In: Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology. pp. 143–151. UIST '98, ACM, New York, NY, USA (1998). https://doi.org/10.1145/288392.288595
5. Nitsche, M., Nürnberger, A.: Quest: Querying complex information by direct manipulation. In: Yamamoto S. (eds) Human Interface and the Management of Information. Information and Interaction Design. HIMI 2013. Lecture Notes in Computer Science **8016** (2006)
6. Russell-Rose, T., Chamberlain, J., Azzopardi, L.: Information retrieval in the workplace: A comparison of professional search practices. Information Processing & Management **54**(6), 1042–1057 (2018). https://doi.org/10.1016/j.ipm.2018.07.003
7. Russell-Rose, T., Gooch, P.: 2dsearch: A visual approach to search strategy formulation. In: Proceedings of DESIRES: Design of Experimental Search Information REtrieval Systems. DESIRES 2018 (2018)
8. Russell-Rose, T., Chamberlain, J.: Real-world expertise retrieval: The information seeking behaviour of recruitment professionals. In: European Conference on Information Retrieval. pp. 669–674. Springer (2016)
9. Russell-Rose, T., Chamberlain, J.: Searching for talent: The information retrieval challenges of recruitment professionals. Business Information Review **33**(1), 40–48 (2016)
10. Tait, J.I.: An introduction to professional search. In: Professional search in the modern world, pp. 1–5. Springer (2014)
11. de Vries, A.P., Alink, W., Cornacchia, R.: Search by strategy. In: Proceedings of the third workshop on Exploiting semantic annotations in information retrieval. pp. 27–28. ACM (2010)
12. Wing Pazer, J.: The importance of the boolean search query in social media monitoring tools. DragonSearch white paper (2013), https://www.dragon360.com/wp-content/uploads/2013/08/social-media-monitoring-tools-boolean-search-query.pdf (retrieved 22-Mar-2018
13. Yi, J.S., Melton, R., Stasko, J., Jacko, J.A.: Dust & magnet: Multivariate information visualization using a magnet metaphor. Information Visualization **4**(4), 239–256 (Oct 2005). https://doi.org/10.1057/palgrave.ivs.9500099