

Dataset Search Queries

UNIVERSITY OF
Southampton



Emilia Kacprzak
@kacprzakemilia

Open Data Portals

219 results

Sort by: Relevance 

Street Lighting Inventory

London Borough of Barnet

The Street Lighting inventory includes details for each street lighting asset maintained by the London Borough of Barnet and comprises information regarding each assets type, its light...

Transport

CSV

Camden Street Lighting

London Borough of Camden

This dataset contains the location of Camden Council owned street lighting across the borough of Camden.

Government

CSV

JSON

RDF

XML

Barnet Street Lighting PFI Contract

London Borough of Barnet

London borough of Barnet Street Lighting PFI Contract for the Installation, Maintenance & Repair of the Borough's street lighting, illuminated signs and traffic bollards assets. More...

Government

PDF

Street Lights

Calderdale Metropolitan Borough Council

Street lights in Calderdale, including coordinates.

Transport

CSV

JSON

XLS

- DCAT metadata
- Human generated metadata
- Search over description

Dataset Search

- Searching for dataset presents challenges
- Not the same user experience as we are used to in web search
- Unique search vertical (as e.g. publication or email search)
- More difficult to explore in comparison to textual sources

Experiment Design

4 Open Data portals:

- Data.gov.uk
- ONS
- Data.gov.ca
- data.gov.au

2 million Portal Queries

500 data request

500 crowd generated queries
(based on selected 50 requests)

Data and Analysis

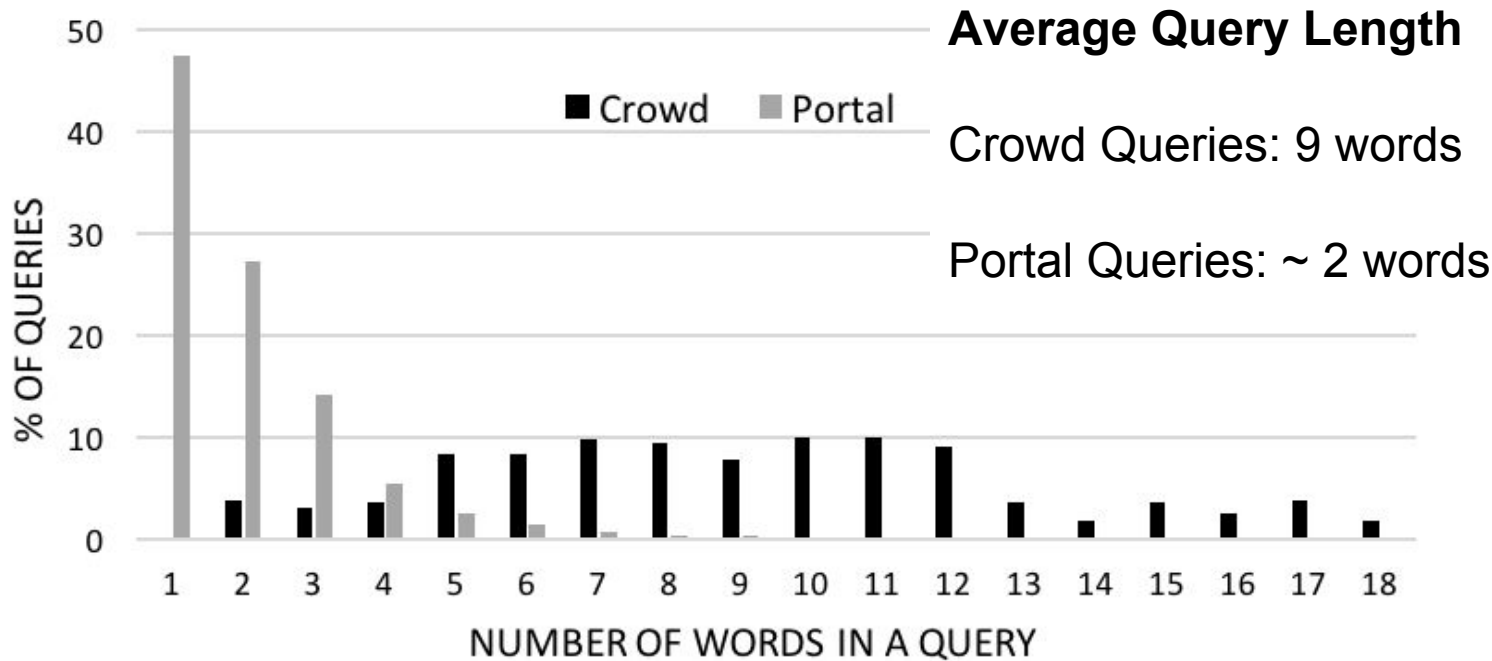
Portal Query:

Yorkshire 1986

Crowd Query:

Annual livestock numbers Yorkshire by parish 1986 onwards

Results - Query Length



Results - Query Types

	Portal Queries	Crowd Queries	
Geospatial	5%	36%	7x more
Temporal	7%	50%	7x more
File and dataset type	6%	50%	8x more
Number	5%	6%	
Only number	<1%	<1%	
Question queries	<1%	9%	7.49% general web search ¹

¹ Michael Bendersky and W. Bruce Croft. 2009. Analysis of Long Queries in a Large Scale Search Log. In Proceedings of the 2009 Workshop on Web Search Click Data. ACM, 8–14. <https://doi.org/10.1145/1507509.1507511>

Summary

- Portals currently do not respond well to longer queries
- Providing faceted search based on geospatial and temporal information
- Longer queries mean more complex information needs
- We need to rethink approaches to index and search through datasets

Thank you!

emilia.kacprzak@theodi.org

[@kacprzakemilia](#)

Data and Analysis

Data Request:

Request annual return data on total numbers of Sheep and Lambs and Cattle and Calves in the following two North Yorkshire parishes from 1986 to the latest available date: for Malham Moor Parish and for Buckden Parish