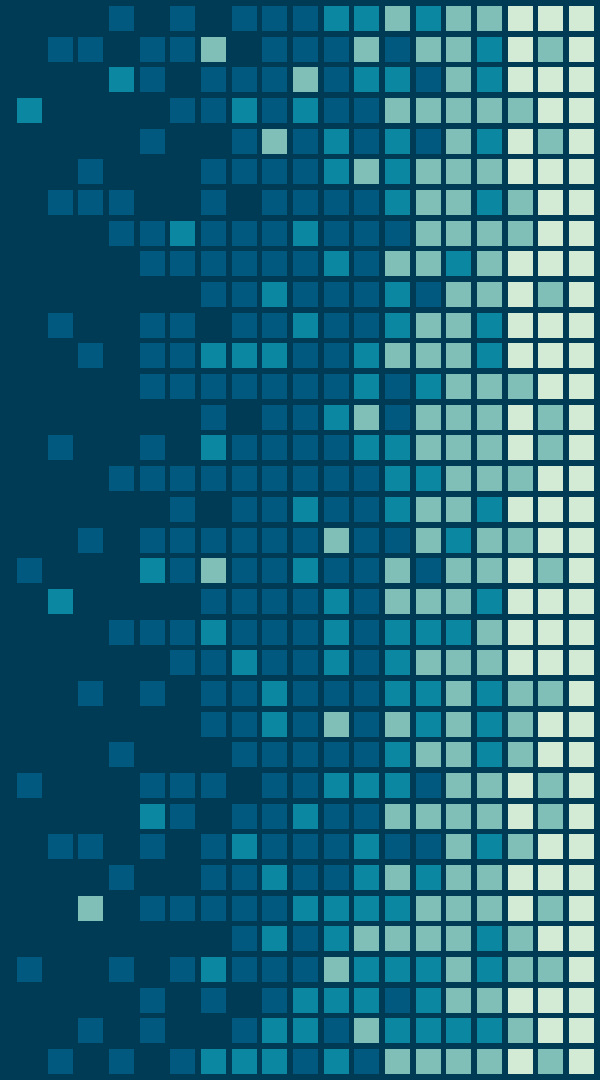


AquaCold

Aggregated Query Understanding And
Construction Over Linked Data

Nick Collis & Ingo Frommholz
University of Bedfordshire



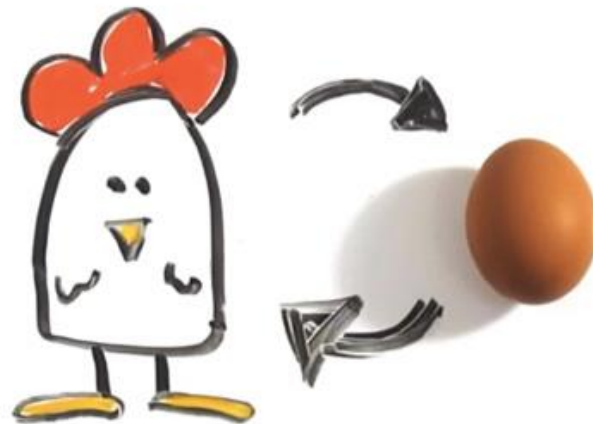
The Linked Data Web today

The Linked Data Web has existed for over 15 years and grown to incorporate **over 149 billion triples**.

Previous problem - Not enough useful linked open data.

Current problem - Now lots of useful linked data but how can we search and explore it effectively?

Then: 'Chicken and egg' problem



Now: *"We have enough chickens, but how can we get enough eggs out of them?"**

* S.Ferre, Expressive and Scalable Query-based Faceted Search over SPARQL Endpoints, 2014

Difficulties with searching linked data

SPARQL remains the most popular method for querying Linked Data. But...

```
PREFIX dbo:
<http://dbpedia.org/ontology/>
PREFIX dbp:
<http://dbpedia.org/resource/>
PREFIX foaf:
<http://xmlns.com/foaf/0.1/>
```

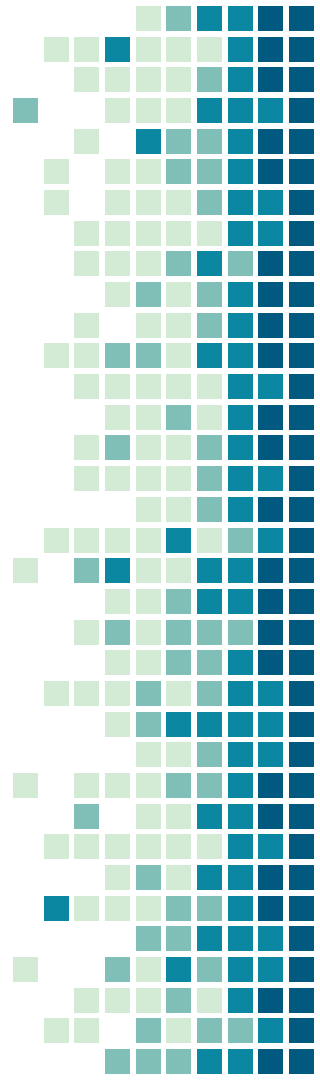
```
SELECT ?name ?bandname where
{
?person foaf:name ?name .
?band dbo:bandMember ?person .
?band dbo:genre dbpedia:Punk_rock .
?band dbp:name ?bandname .
}
```

SPARQL syntax is unintuitive for non technical users.

The SPARQL endpoint must be known in order to write queries.

Users must understand the **RDF model, and ontology terms**

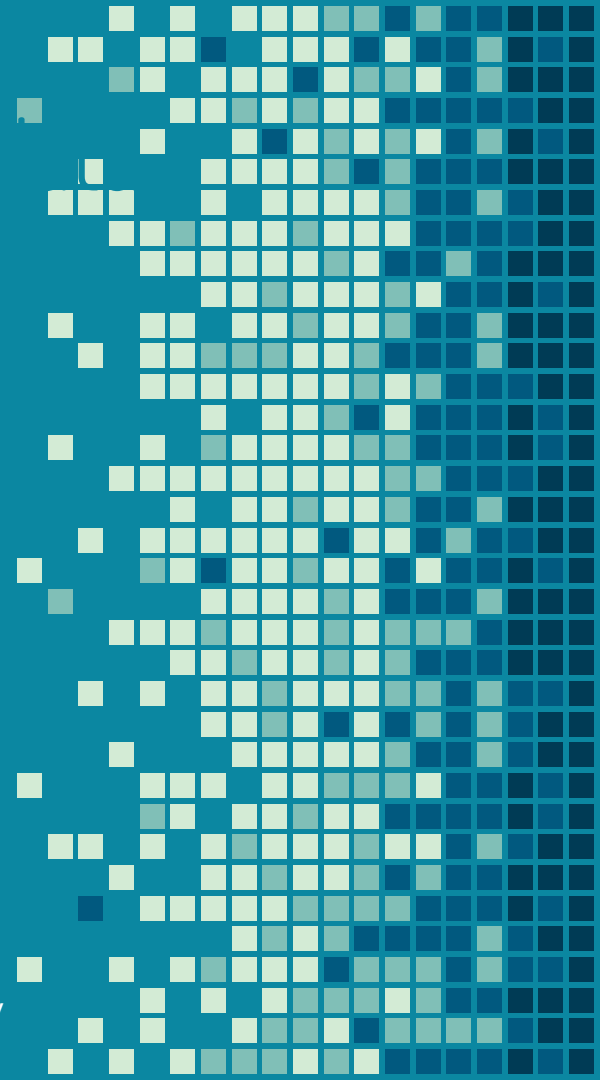
Many instances of ambiguity in data (eg :Place vs PopulatedPlace)





“...the lack of technical knowledge and an understanding of the intricacies of the semantic technology stack limits users in their ability to interpret and make use of the Web of Data...”

...the key solution is to visualise Linked Data in a coherent and legible manner, allowing non-domain and non-technical audiences to obtain an understanding of its structure, and therefore implicitly compose queries, identify links between resources and intuitively discover new pieces of information”



Goals

Make linked data search more usable:

- **Abstract the complexity** of SPARQL whilst retaining expressivity
- Improve the **discoverability** of endpoints and ontologies
- Provide a measurement of **result accuracy**
- Provide a way of identifying **ambiguous** labels
- Build a **scalable** system, able to cope with vast amounts of data



Approaches to improving LD search usability

Graphical Query Builders

- + More readable than SPARQL
- Less intuitive than NL
- Lacks expressivity

Controlled Natural Language

- + Readable, intuitive, robust
- Inflexible

Uncontrolled Natural Language

- + Highly intuitive when working
- Less accurate results, particularly for complex queries

1. Ask a question

Start by asking a question in natural language and watch the query generated:

Question: Which books did Stephen King write?

This query was generated for the question "Which books did Stephen King write?"

SPARQL

EDIT

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX quepy: <http://www.machinalis.com/quepy#>
PREFIX dbpedia: <http://dbpedia.org/ontology/>
PREFIX dbpprop: <http://dbpedia.org/property/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
```

```
SELECT DISTINCT ?x2 WHERE {
  ?x0 rdf:type dbpedia-owl:Book.
  ?x0 dbpedia-owl:author ?x1.
  ?x0 foaf:name ?x2.
```

Quepy

The Aquacold interface

Tanks used in World War 2

1

search

URI	URILabel	type	used in war	
				Renault
http://dbpedia.org/resource/Hotchkiss_H35	Hotchkiss H35	Light tank	World War II, Israeli War o...	▶ manufacturers (11) > Renault ▶ manufacturer (8) > Renault
http://dbpedia.org/resource/Renault_FT	Renault FT	Light tank	1948 Arab–Israeli War, C...	
http://dbpedia.org/resource/Panzer_35(t)	Panzer 35(t)	Light tank	World War II	
http://dbpedia.org/resource/Vickers_6-Ton	Vickers 6-Ton	Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/BT_tank	BT tank	Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/T-60_tank	T-60 tank	Light tank	World War II	

↑
1891
↓

save query

1 Search & Labelling box – Users type NL query here and also label the results grid using NL

The Aquacold interface

Tanks used in World War 2

1

search

2

URI	URILabel	type	used in war	
				Renault
http://dbpedia.org/resource/Hotchkiss_H35	Hotchkiss H35	Light tank	World War II, Israeli War o...	<ul style="list-style-type: none">manufacturers (11) > Renaultmanufacturer (8) > Renault
http://dbpedia.org/resource/Renault_FT	Renault FT	Light tank	1948 Arab–Israeli War, C...	
http://dbpedia.org/resource/Panzer_35(t)	Panzer 35(t)	Light tank	World War II	
http://dbpedia.org/resource/Vickers_6-Ton	Vickers 6-Ton	Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/BT_tank	BT tank	Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/T-60_tank	T-60 tank	Light tank	World War II	

↑
1891
↓

save query

2 Search button – On clicking this, the results grid will be refreshed with the results of the query.

The Aquacold interface

Tanks used in World War 2

1

search

2

URI	URILabel	3	type	used in war	
					Renault
http://dbpedia.org/resource/Hotchkiss_H35	Hotchkiss H35		Light tank	World War II, Israeli War o...	<ul style="list-style-type: none">manufacturers (11) > Renaultmanufacturer (8) > Renault
http://dbpedia.org/resource/Renault_FT	Renault FT		Light tank	1948 Arab–Israeli War, C...	
http://dbpedia.org/resource/Panzer_35(t)	Panzer 35(t)		Light tank	World War II	
http://dbpedia.org/resource/Vickers_6-Ton	Vickers 6-Ton		Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/BT_tank	BT tank		Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/T-60_tank	T-60 tank		Light tank	World War II	

↑
1891
↓

save query

3 Property filters – Used to search for properties of an object, such as type, age, weight, etc..

The Aquacold interface

Tanks used in World War 2

1

search

2

URI	URILabel	type	used in war	
				Renault
http://dbpedia.org/resource/Hotchkiss_H35	Hotchkiss H35	Light tank	World War II, Israeli War o...	▶ manufacturers (11) > Renault ▶ manufacturer (8) > Renault
http://dbpedia.org/resource/Renault_FT	Renault FT	Light tank	1948 Arab–Israeli War, C...	
http://dbpedia.org/resource/Panzer_35(t)	Panzer 35(t)	Light tank	World War II	
http://dbpedia.org/resource/Vickers_6-Ton	Vickers 6-Ton	Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/BT_tank	BT tank	Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/T-60_tank	T-60 tank	Light tank	World War II	

4

↑
1891
↓

save query

4

Subject & property filters – Used to search for a property / value pair.

The Aquacold interface

Tanks used in World War 2

1

search

2

URI	URILabel	type	used in war	
				Renault
http://dbpedia.org/resource/Hotchkiss_H35	Hotchkiss H35	Light tank	World War II, Israeli War o...	▶ manufacturers (11) > Renault ▶ manufacturer (8) > Renault
http://dbpedia.org/resource/Renault_FT	Renault FT	Light tank	1948 Arab–Israeli War, C...	
http://dbpedia.org/resource/Panzer_35	Panzer 35(t)	Light tank	World War II	
http://dbpedia.org/resource/Vickers_6-Ton	Vickers 6-Ton	Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/BT_tank	BT tank	Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/T-60_tank	T-60 tank	Light tank	World War II	

Renault

▶ manufacturers (11) > Renault

▶ manufacturer (8) > Renault

5

↑
1891
↓

save query

5 Results grid – Lists the results for this query. These can be manipulated using the filters.

The Aquacold interface

Tanks used in World War 2

1

search

2

URI	URILabel	type	used in war	
				Renault
http://dbpedia.org/resource/Hotchkiss_H35	Hotchkiss H35	Light tank	World War II, Israeli War o...	▶ manufacturers (11) > Renault ▶ manufacturer (8) > Renault
http://dbpedia.org/resource/Renault_FT	Renault FT	Light tank	1948 Arab–Israeli War, C...	
http://dbpedia.org/resource/Panzer_35(t)	Panzer 35(t)	Light tank	World War II	
http://dbpedia.org/resource/Vickers_6-Ton	Vickers 6-Ton	Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/BT_tank	BT tank	Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/T-60_tank	T-60 tank	Light tank	World War II	

6

↑
1891
↓

save query

6 Results grid voting – Users vote on the accuracy of the results and label using these controls.

The Aquacold interface

Tanks used in World War 2

1

search

2

URI	URILabel	type	used in war	
				Renault
http://dbpedia.org/resource/Hotchkiss_H35	Hotchkiss H35	Light tank	World War II, Israeli War o...	<ul style="list-style-type: none">manufacturers (11) > Renaultmanufacturer (8) > Renault
http://dbpedia.org/resource/Renault_FT	Renault FT	Light tank	1948 Arab–Israeli War, C...	
http://dbpedia.org/resource/Panzer_35	Panzer 35(t)	Light tank	World War II	
http://dbpedia.org/resource/Vickers_6-Ton	Vickers 6-Ton	Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/BT_tank	BT tank	Light tank	Winter War, World War II, ...	
http://dbpedia.org/resource/T-60_tank	T-60 tank	Light tank	World War II	

6

1891

7

save query

7

Save query button – This saves the query to the database with the label defined in box 1.

AquaCold - Querying

Films starring Tom

URI	Property
▶ Tom Arnold	
▶ Tom Hardy	
▶ Tom Hiddleston	
▶ Tommy Lee Jones	

```
SELECT ?x ?y ?z
WHERE {
  ?z rdf:type <dbpedia-owl:Film> .
  ?z <dbpedia-owl:starring> ?y .
  ?z <dbpedia-owl:starring> ?x .
  ?y <rdfs:label> ?variable1 .
  ?y bif:contains 'Tom'
}
```

SPARQL QUERY TEMPLATE
RETRIEVES MATCHING ENTITIES

- Tom Arnold
- Tom Cruise
- Tom Hardy
- Tom Hiddleston
- Tommy Lee Jones
- Tom Waits

Autocomplete suggestions are based on labels written by other users

AquaCold - Building result grids

```
SELECT ?property ?y
WHERE {
  ?x ?property ?y .
  ?y rdf:label ?z .
  FILTER regex(?z, "Al Pac", "i")
}
```

URI	URILabel	property	
		Al Pac	
		▶ film director (4) > Al Pacino	
		▶ portrayer (3) > Al Pacino	
		▶ starring (46) > Al Pacino	
		▶ writer (3) > Al Pacino	

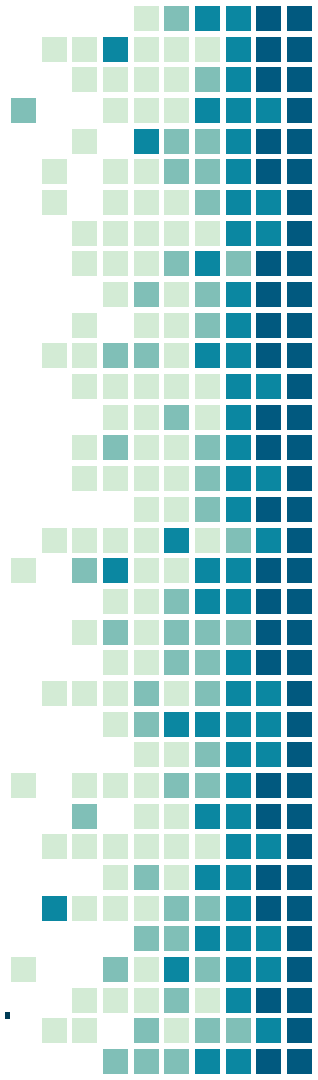
URI	URILabel	▲starring
http://dbpedi...	...And Justice for All (film)	John Forsythe, John Forsythe, Lee Strasberg, Lee Strasberg, Al Pacino
http://dbpedi...	Author! Author! (film)	Tuesday Weld, Dyan Cannon, Bob Dishy, Al Pacino, Alan King (cameo)
http://dbpedi...	Glengarry Glen Ross (film)	Jack Lemmon, Kevin Spacey, Alan Arkin, Jonathan Pryce, Al Pacino, Al Pacino

Entity autocomplete is carried out using basic regex string matching

AquaCold - Labelling Linked Data Sets

Horror fiction books by	search		author	literary genre
▶ Horror fiction books by Rob Hood				
▶ Horror fiction books by Hugh B. Cave				
▶ Horror fiction books by R. Chetwynd-Hayes				
▶ Horror fiction books by Robert M. Price				
▶ Horror fiction books by Brian Evenson	rg/...	Children of the Corn	Stephen King	Horror fiction
▶ Horror fiction books by Kim Harrison	rg/...	Willa (short story)	Stephen King	Horror fiction
▶ Horror fiction books by Stephenie Meyer	rg/...	The Ledge (short story)	Stephen King	Horror fiction
▶ Horror fiction books by August Derleth	rg/...	Crouch End (short story)	Stephen King	Horror fictio...
▶ Horror fiction books by Kelley Armstrong	rg/...	Quitters, Inc.	Stephen King	Horror fiction
▶ Horror fiction books by Edward Lee (writer)	rg/...	Survivor Type	Stephen King	Horror fiction
▶ Horror fiction books by Arthur Machen	rg/...	The Night Flier	Stephen King	Horror fiction
▶ Horror fiction books by Bram Stoker	rg/...	1922 (novella)	Stephen King	Horror fiction
▶ Horror fiction books by Algernon Blackwood	rg/...	Fair Extension	Stephen King	Comedy hor...
▶ Horror fiction books by Jeff VanderMeer	rg/...	Night Shift (short story collection)	Stephen King	Horror fiction
▶ Horror fiction books by Jack Ketchum	rg/...	Rainy Season (short story)	Stephen King	Horror fiction
▶ Horror fiction books by Rod Serling				
▶ Horror fiction books by Anthony Shaffer (writer)				
▶ Horror fiction books by M. T. Anderson				
▶ Horror fiction books by Dean Koontz				

Result grids can be labelled with NL, guided by labels used by others.



AquaCold - Voting

URI	URILabel	author	literary genre
http://dbpedi...	Children of the Corn	Stephen King	Horror fiction
http://dbpedi...	Willa (short story)	Stephen King	Horror fiction
http://dbpedi...	The Ledge (short story)	Stephen King	Horror fiction
http://dbpedi...	Crouch End (short story)	Stephen King	Horror fiction, ...



URI	URILabel	literary genre	author
http://dbpedi...	Your Heart Belongs to Me (novel)	Gothic fiction, Sus...	Dean Koontz
http://dbpedi...	Invasion (Koontz novel)	Suspense, Horror f...	Dean Koontz
http://dbpedi...	77 Shadow Street	Science fiction, Thr...	Dean Koontz
http://dbpedi...	Mr. Murder	Horror fiction	Dean Koontz



User Score

  **+1,513**

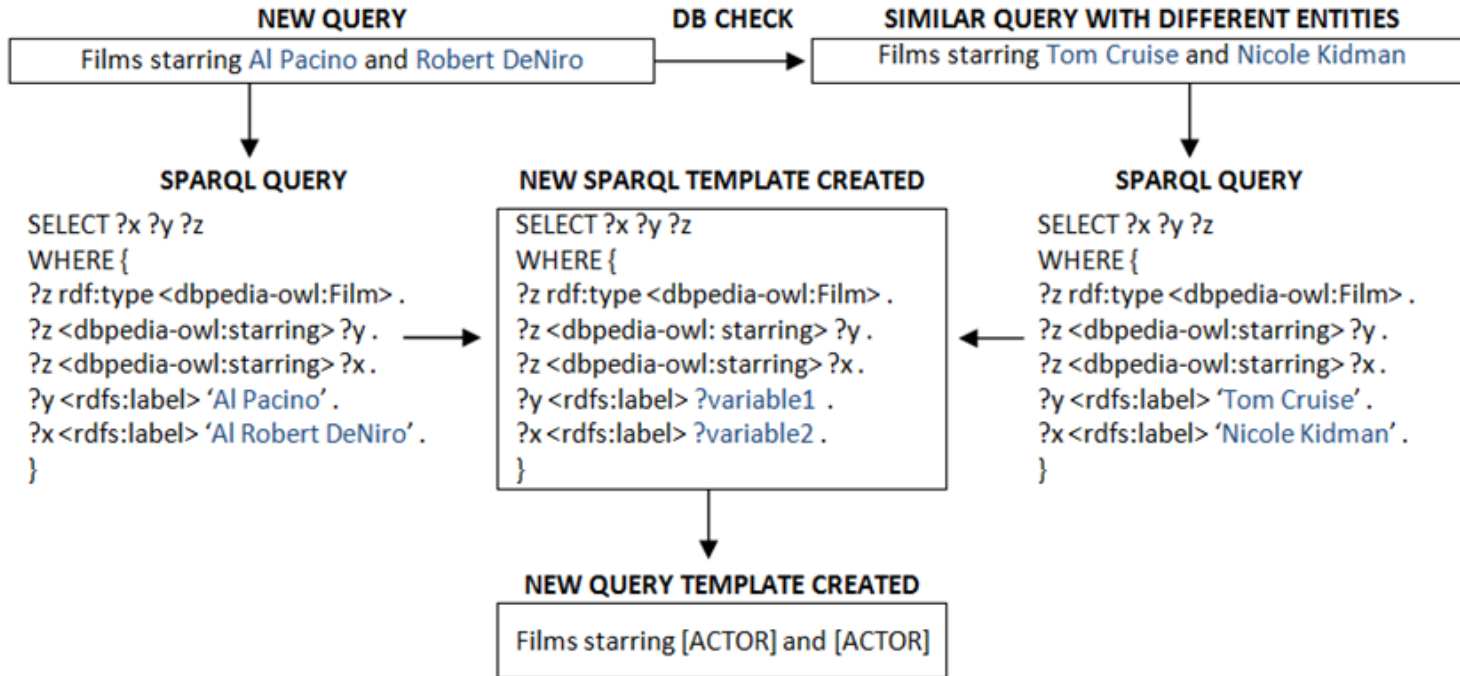
User Score

  **-2,983**

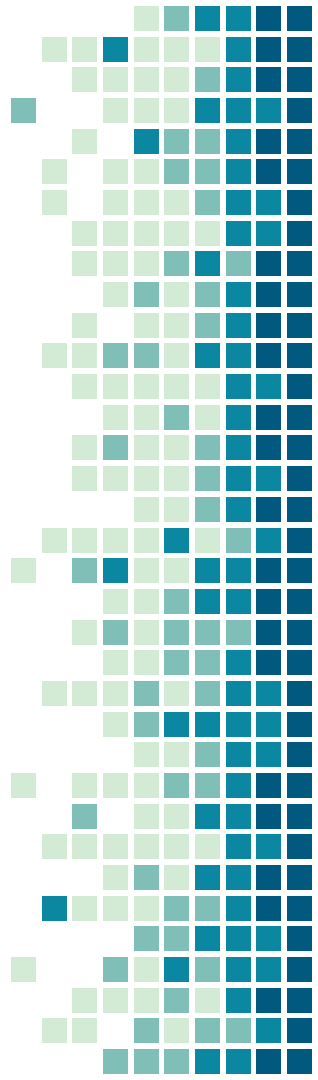
Horror Fiction books by Stephen King
▶ Horror Fiction books by Stephen King (+1,513 votes)
▶ Horror Fiction books by Stephen King (+952 votes)
▶ Horror Fiction books by Stephen King (+5 votes)
▶ Horror Fiction books by Stephen King (-20 votes)
▶ Horror Fiction books by Stephen King (-752 votes)
▶ Horror Fiction books by Stephen King (-786 votes)
▶ Horror Fiction books by Stephen King (-1,013 votes)
▶ Horror Fiction books by Stephen King (-2,983 votes)
▶ Horror Fiction books by Stephen King (-3,982 votes)

Uses crowdsourced voting on labels to surface high quality results

AquaCold - Templating



Creates templates based on labels to seed multiple label variants.



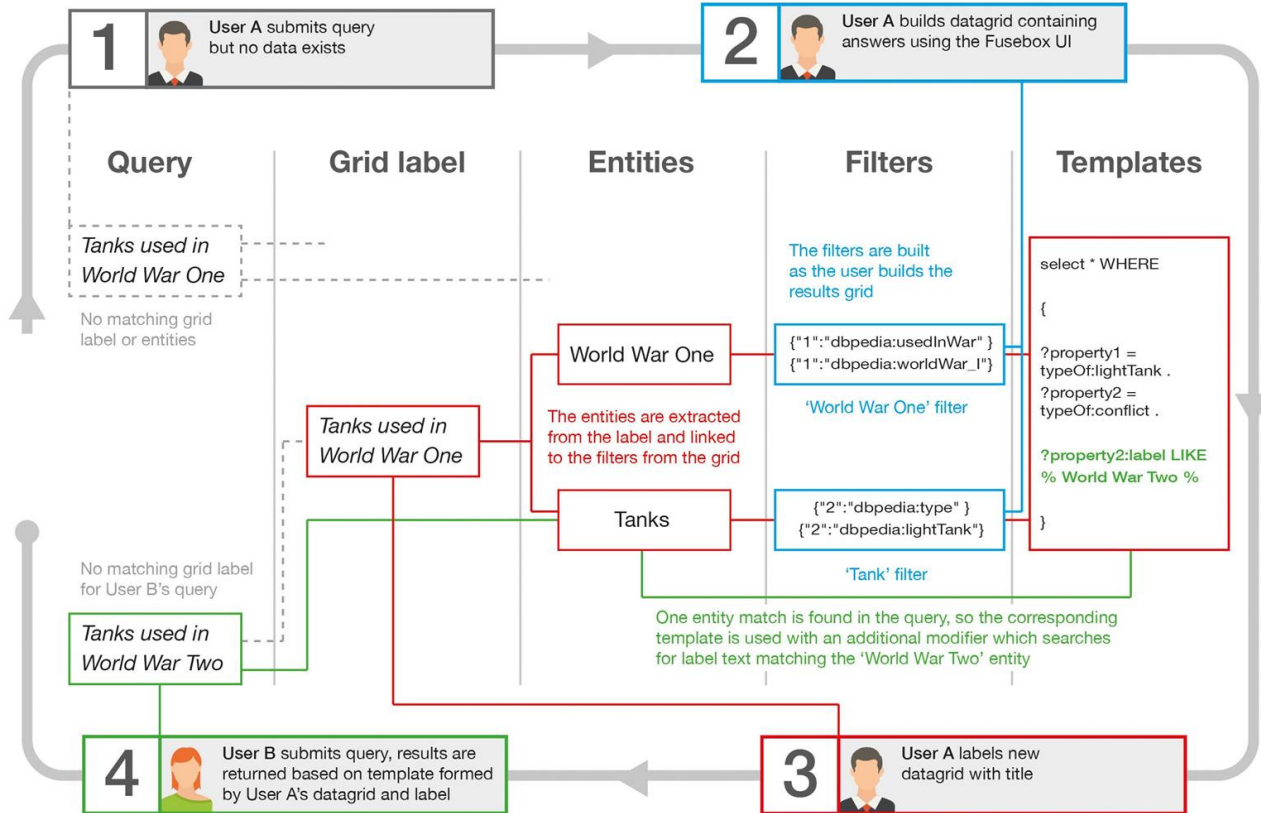
Demonstration

[Video 1](#)

[Video 2](#)

[Video 3](#)

[Video 4](#)



User A
Looking for
Tanks used in World War One



User B
Looking for
Tanks used in World War Two

How AquaCold allows usable Linked Data Search

Abstract SPARQL complexity retaining expressivity

Natural language input abstracts complexity of SPARQL for querying
Faceted results grid retains some SPARQL expressivity for exploration

Improve discoverability of *endpoints* and *ontologies*

Users discover endpoints, build result grids and label for other users

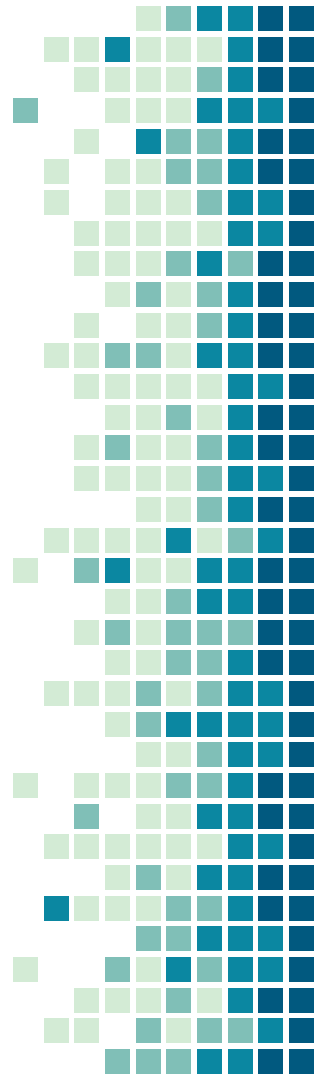
Providing a measurement of accuracy for results

Crowdsourced voting tools surface the most accurate results

Enabling disambiguation of result labels

Multiple identical labels are allowed (attached to different result grids)

Voting mechanism surfacing the most relevant



More benefits of AquaCold approach

Natural language queries for linked data result sets.

- Highly intuitive for non technical users. Allows synonyms, slang etc.

Guided labelling assistance.

- Displays similar existing labels to limit label ambiguity.

Faceted interface for composing result sets.

- Hides SPARQL complexity. No need to know the endpoint details.

Voting mechanism for labels and result sets.

- No need to know the endpoint, ontology terms or URIs.

Template label generation.

- Ameliorates cold start problem



Limitations of AquaCold approach

Basic voting system open to misuse

- Simple +1, -1 voting could be easily manipulated.

Grid based interface suits some queries more than others

- Queries with many links between nodes. Eg FOAF less suited

Could lead to issues with ambiguity in query labels.

- Disparity and ambiguity in LD could transfer to AquaCold

System is very early in development

- Many features not yet present, eg: AND, NOT, node linking

Lacks expressivity compared to SPARQL

- Cannot match the expressivity of dedicated query language



Evaluation planned for early 2018

Evaluation will use questions from the QALD challenge

Established benchmark used by other NL systems inc. SPARKLIS, TBSL

Measured against similar NL LD search systems

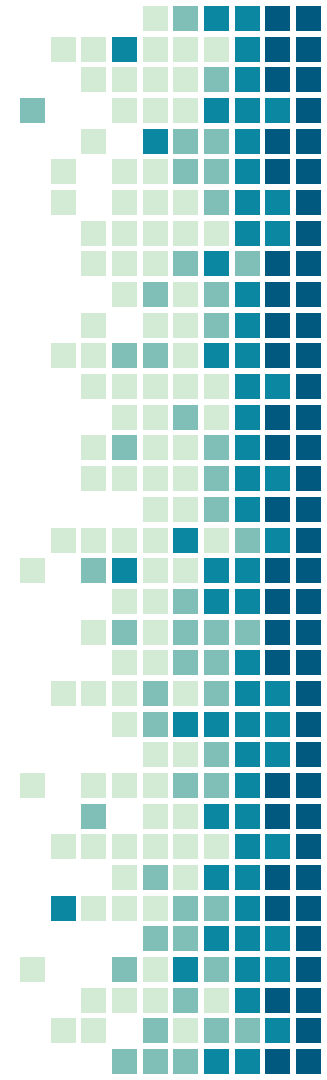
Including SPARKLIS, CrowdQ, Quepy, Ginseng, Semantic Crystal

Key challenge - parity with other systems

Allow accurate comparisons with other systems taking into account features that are unique to AquaCold

Test groups to have a range of experience levels

Will measure **Expressivity, Scalability and Usability**



Future work

Hypothesis

human labelled result sets and crowdsourced quality control will result in higher precision compared to systems which use programmatic translation of natural language queries to retrieve linked data.

Enhance crowdsourcing & voting system:

Investigate literature for enabling effective crowdsourced voting

Build more advanced filters

NOT, AND, numeric < >, date operators, fuzzy regex matching....

Enhance entity detection

Investigate entity detection tools such as IBM Watson or Word2Vec



THANKS!

Any questions?

You can find me at:
nwcphd@gmail.com

