Combining Flexible Queries and Knowledge Anchors to facilitate the exploration of Knowledge Graphs

Alex Poulovassilis[1], Marwan Al-Tawil[2], Riccardo Frosini[1], Mirko Dimartino[1], Vania Dimitrova[2]

[1]Knowledge Lab, Birkbeck University of London
[2]School of Computing, University of Leeds

IESD 2016 @ ISWC 2016

Outline of the talk

1. Introduction and Motivation
2. Case Study – L4All Ontology and Dataset
3. Knowledge Anchor Derivation
4. Extending Flexible Querying with Knowledge Anchors
5. Conclusions and Future work

1. Introduction and Motivation

- Increasing volumes of graph-structured data arising from many application areas, e.g. RDF linked data on the web
- Applications seek to take advantage of such knowledge graphs
  - in domains such as web information retrieval, formal and informal learning, health informatics, entertainment, cultural heritage preservation
- Volumes, complexity and heterogeneity of the data means that users are unlikely to be familiar with its full structure and content
- Hence need to be assisted by intelligent tools that support interactive exploration of the knowledge graph
- Recent work has proposed techniques for *flexible querying* to find paths through knowledge graphs
- Amongst other work, we have extended SPARQL 1.1 with *approximation* and *relaxation* operations, calling the resulting language **SPARQL^{AR}**
  - See Cali et al ODBASE 2014, Frosini et al Sem. Web J. 2016 forthcoming

1. Introduction and Motivation

- Features of our flexible querying approach include that:
  - users' queries do not have to match exactly the data structures being queried
  - the query system can automatically make changes to a query so as to help the user find relevant information
  - answers to queries are returned in ranked order, in increasing 'distance' from the original query
    - this distance is the sum of the costs of all the approximation or relaxation operations that have been applied to the original query in order to obtain the answer
    - for the purposes of this paper, we assume that all such operations have a cost of 1, but in general these can be application-defined or user-defined

Introduction and Motivation

- We support two kinds of flexible querying in **SPARQL^{AR}** :
  - query *relaxation*
    - applies a *relaxation operation* to the query, e.g. replacing a class by a superclass, or replacing a property by a superproperty
    - returns *additional* answers compared to the exact form of the query
  - query *approximation*
    - applies an *edit operation* to the query e.g. the insertion, deletion or substitution of an edge label
    - returns *different* answers compared to the exact form of the query

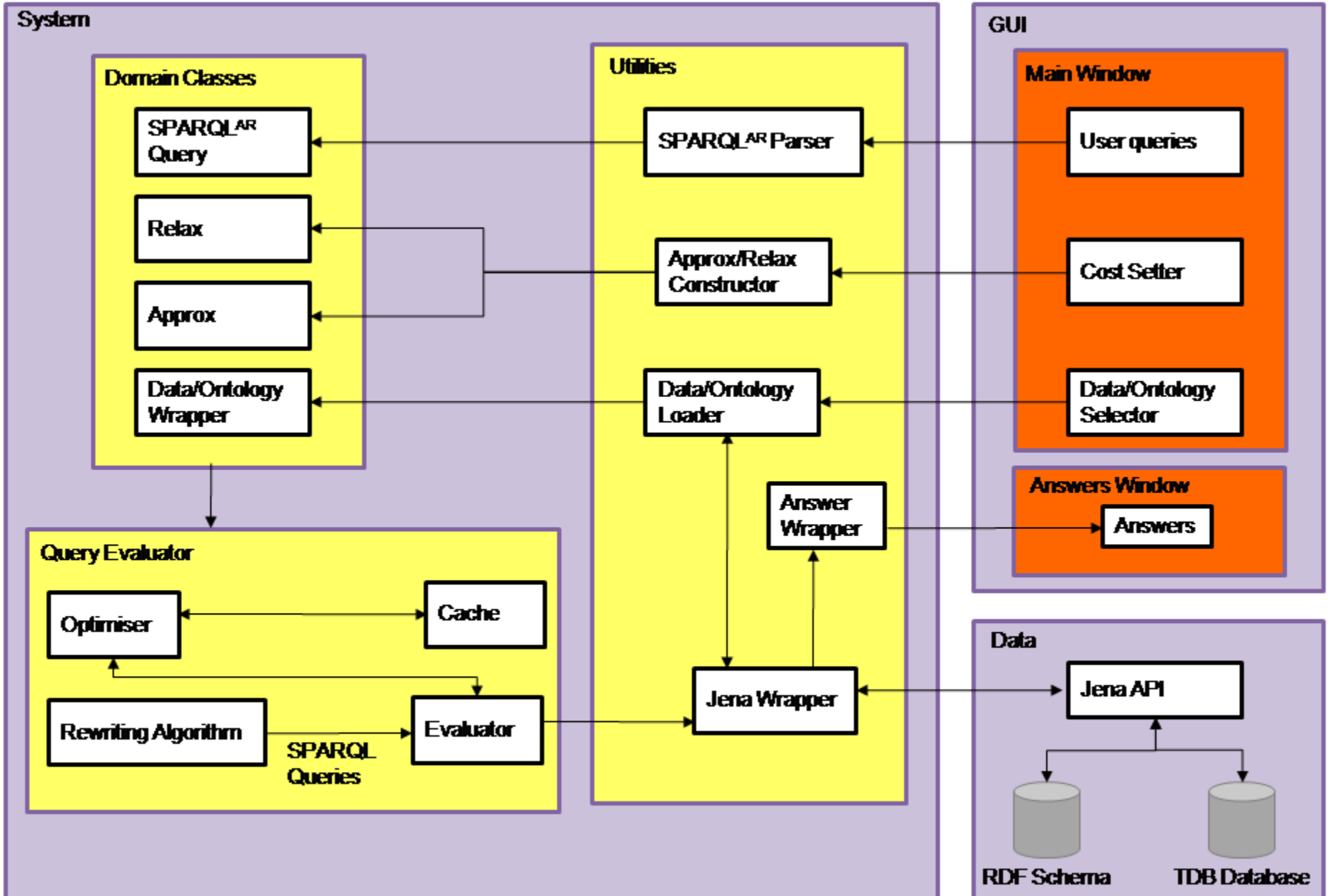
Introduction and Motivation

- Example SPARQL^{AR} query:

```
SELECT ?WorkEp ?Occ
WHERE {?EdEp rdf:type <http://www.L4All.com/University_Episode>.
      RELAX ( ?EdEp <http://www.L4All.com/qualif>/rdf:type
                  <http://www.L4All.com/Information_Systems> ).
      APPROX( ?EdEp <http://www.L4All.com/prereq> ?WorkEp ).
      ?WorkEp  rdf:type  <http://www.L4All.com/Work_Episode>.
      ?WorkEp  <http://www.L4All.com/job>/rdf:type ?Occ}
```

- Of course, this is not what an end-user would enter, but it could be the SPARQL^{AR} query generated from the user's interaction with a graphical user interface, e.g. such as that described in Poulovassilis et al IEEE TLT 2012

SPARQL^{AR} system architecture



Introduction and Motivation

- Supporting such flexible query processing over knowledge graphs brings several benefits:

  - automatic correction of users' erroneous queries;
  - finding additional relevant answers that the user may be unaware of;
  - generating new queries which may return unexpected results and bring new insights

Introduction and Motivation

- Although flexible query processing allows broadening a user's perspective of the knowledge domain, it can return a large number of results, all at the same `distance' away from the user's original query
- Therefore, a key challenge is how to *facilitate users' meaning making from flexible query results*
- Recent work has investigated supporting users' sense-making and knowledge expansion via *interactive nudges*:
  - detecting which entities a user may be familiar with (e.g. by analysing interaction logs) is a computationally intensive task
  - moreover, when the user has had limited interaction with the system, there is the well-known 'cold start' problem
  - other ways are needed for automatically identifying which entities may be close to the users' cognitive structures, and hence may offer good ***knowledge anchors*** for information exploration.

Introduction and Motivation

- Recent work has proposed an approach to identifying knowledge anchors that adopts the Cognitive Science notion of *basic-level objects* in domain taxonomies
  - See Al-Tawil et al ACM Conf. on Hypertext and Social Media, 2016

- This work has developed a formal framework for identifying knowledge anchors in knowledge graphs, applying two complementary approaches:
  - *distinctiveness* and *homogeneity*

- Distinctiveness metrics identify the most differentiated domain categories, whose attributes are associated with the category members but not with members of other categories
- Homogeneity metrics identify categories whose members share many attributes.

Identifying Knowledge Anchors (HT2016)

UNIVERSITY OF LEEDS

Distinctiveness

Input: $DG = \langle V, E, P \rangle, e \in E$

1. **for all** $v \in \{C\}$ **do**
2. $V' :=$ the set of all $v' : v' \subseteq v$
3. **for all** $v'_e : \exists \langle v'_e, e, v' \rangle$ **do**
4. $N_e :=$ set of all $\langle v'_e, e, v' \rangle : v' \in V'$
5. $M_e :=$ set of all $\langle v'_e, e, v_a \rangle : v_a \in V$
6. $AV_{v'_e} := |N_e| / |M_e|$
7. $CAC_{v'_e} := (|N_e| / |M_e|) \cdot (|N_e| / |V'|)$
8. $CU_{v'_e} := (|N_e| / |V'|)^2 - (|M_e| / |V|)^2$
9. $AV_v := AV_v + AV_{v'_e}$
10. $CAC_v := CAC_v + CAC_{v'_e}$
11. $CU_v := CU_v + CU_{v'_e}$
12. **end for**
13. **end for**

Output: AV_v, CAC_v, CU_v for all $v \in \{T \cup C\}$

Homogeneity

Input: $DG = \langle V, E, P \rangle, e \in E$

1. **for all** $v \in \{C\}$ **do**
2. $V' :=$ the set of all $v' : v' \subseteq v$
3. **for all** $(v', v'') : v' \in V' \wedge v'' \in V'$ **do**
4. $V'_e := \{v'_e : \exists \langle v'_e, e, v' \rangle\}$
5. $V''_e := \{v''_e : \exists \langle v''_e, e, v'' \rangle\}$
6. $I := V'_e \cap V''_e$
7. $U := V'_e \cup V''_e$
8. $CN_{v',v''} := |I|$
9. $Jac_{v',v''} := |I| / |U|$
10. $Cos_{v',v''} := |I| / (\sqrt{|V'_e|} \cdot \sqrt{|V''_e|})$
11. $CN_v = CN_v + CN_{v',v''}$
12. $Jac_v = Jac_v + Jac_{v',v''}$
13. $Cos_v = Cos_v + Cos_{v',v''}$
14. **end for**
15. $CN_v = CN_v / (|V'| . (|V'| - 1) / 2)$
16. $Jac_v = Jac_v / (|V'| . (|V'| - 1) / 2)$
17. $Cos_v = Cos_v / (|V'| . (|V'| - 1) / 2)$
18. **end for**

Output: CN_v, Jac_v, Cos_v for all $v \in \{C\}$

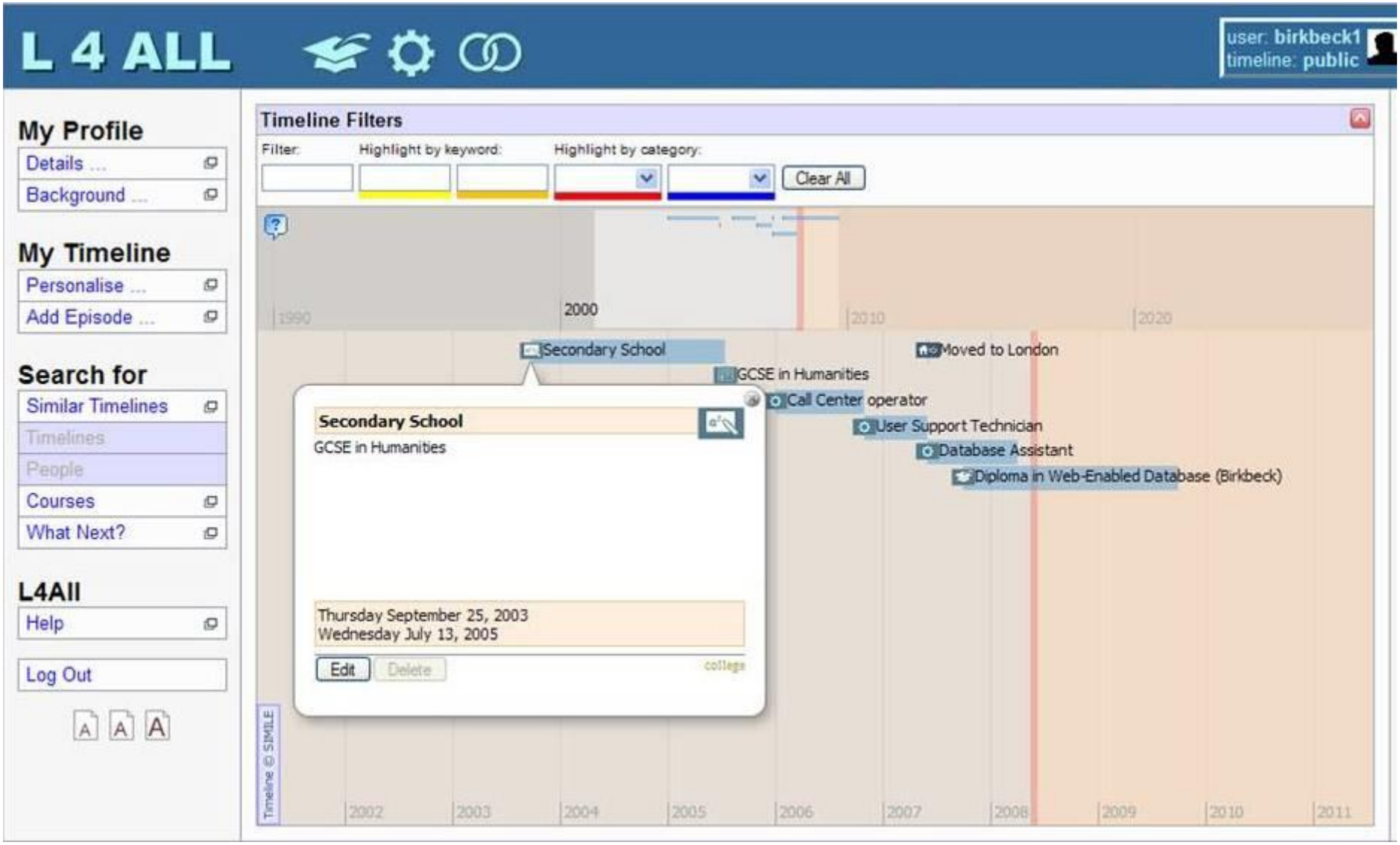
Introduction and Motivation

- We draw on these two recent strands of work (namely, flexible querying of graph-structured data and identification of good anchors for knowledge graph exploration) in order to support users in incrementally querying, exploring and learning from large, complex knowledge graphs.

- We illustrate this integrative approach through a case study in exploring career options.

2. Case Study – L4All Ontology and Dataset

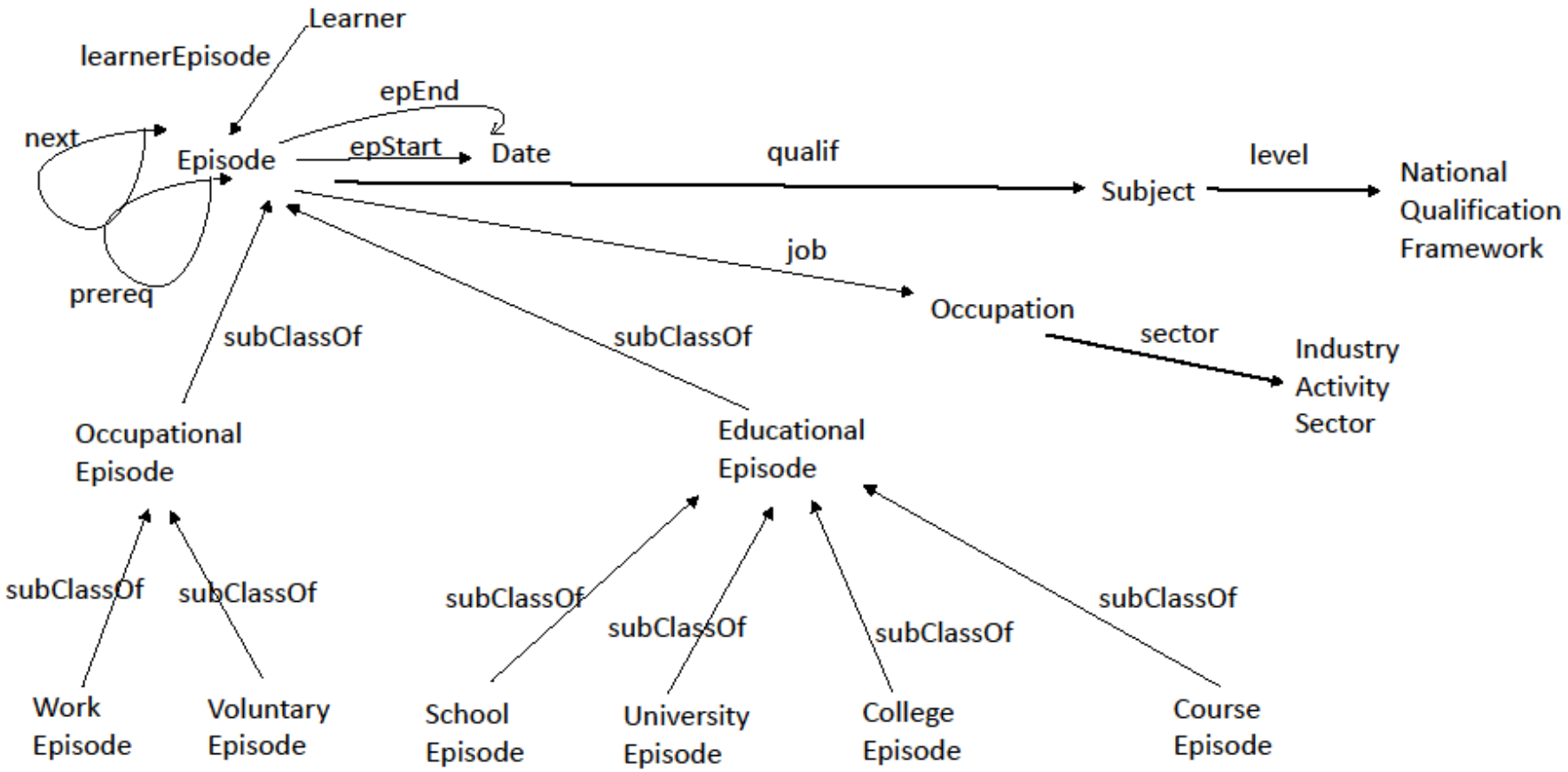
- The L4All project aimed to provide lifelong learners with access to information and resources that would support them in exploring learning and career opportunities and in planning and reflecting on their learning
- It brought together experts from lifelong learning and careers guidance, content providers, and groups of students and tutors
- The L4All pilot system allowed users to record their past learning, work and life experiences within a 'timeline'
- Users' timelines are encoded in the form of RDF/S
- Some types of episode can be annotated by the user with a primary and possibly a secondary classification
- These classifications are drawn from standard Occupational and Educational taxonomies of the UK Office for National Statistics (see later)


L 4 ALL
user: birkbeck1
timeline: public
My Profile
Details ...
Background ...
My Timeline
Personalise ...
Add Episode ...
Search for
Similar Timelines
Timelines
People
Courses
What Next?
L4All
Help
Log Out
Timeline Filters
Filter:
Highlight by keyword:
Highlight by category:
Clear All
1990
2000
2010
2020
Secondary School
GCSE in Humanities
Moved to London
Call Center operator
User Support Technician
Database Assistant
Diploma in Web-Enabled Database (Birkbeck)
Secondary School
GCSE in Humanities
Thursday September 25, 2003
Wednesday July 13, 2005
Edit
Delete
college
Timeline © SIMILE
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011

Case Study – L4All Ontology and Dataset

- Users are able to search over the timelines of other learners and alumni (for those timelines that have been made public by their owner)
  - gives a repertoire of learning and work possibilities that they may not have otherwise considered,
  - allows sharing of successful learning pathways,
  - presenting successful learners as role models to inspire confidence and a sense of opportunity.
- In the original pilot system, *similarity measures* were used for comparing users' timelines
- However, this had a number of drawbacks, largely arising from the rigidity of the similarity matching algorithms employed
- Led to further work that investigated the use of *flexible query processing* techniques based on query approximation and relaxation to support users' search over the timeline data

Case Study – L4All Ontology and Dataset



Case Study – L4All Ontology and Dataset

Class hierarchy	Depth	No. of classes
Episode	3	9
Subject	3	161
Occupation	4	465
National Qualification Framework	3	36
Industry Activity Sector	3	22
	Total:	693

3. Knowledge Anchor Derivation

- We generated the Knowledge Anchors from the L4All ontology combined with an RDF dataset comprising data relating to 1700 timelines (called dataset L4All2 in the paper)
- Knowledge Anchors aim to represent *familiar and highly inclusive entities* in the graph from which links to new knowledge can be made
- This new knowledge can take meaning by becoming linked to existing concepts within the user's cognitive structures.
- We considered two types of relationships:
  - *hierarchical relationships* denoting membership between the subject and object of the corresponding RDF triples (for our case study here, rdfs:subClassOf and rdf:type);
  - *domain-specific relationships* are properties other than the hierarchical relationships

Knowledge Anchor Derivation

- We adopted two complementary groups of metrics to identify knowledge anchors (see Al-Tawil, Dimitrova, Thakker, Bennet ACM Conf. on Hypertext and Social Media, 2016 for a detailed description of the metrics and algorithms):
  - *Distinctiveness metrics* identify the most differentiated categories, whose attributes are associated with the category members but not with members of other categories. We use three distinctiveness metrics: Attribute Validity (AV), Category-Attribute Collocation (CAC), Category Utility (CU).
  - *Homogeneity metrics* identify categories whose members share many attributes. We use three set-based similarity metrics: Common Neighbors (CN), Jaccard (Jac), Cosine (Cos).

Identifying Knowledge Anchors (HT2016)

UNIVERSITY OF LEEDS

Distinctiveness

Input: $DG = \langle V, E, P \rangle, e \in E$

1. **for all** $v \in \{C\}$ **do**
2. $V' :=$ the set of all $v' : v' \subseteq v$
3. **for all** $v'_e : \exists \langle v'_e, e, v' \rangle$ **do**
4. $N_e :=$ set of all $\langle v'_e, e, v' \rangle : v' \in V'$
5. $M_e :=$ set of all $\langle v'_e, e, v_a \rangle : v_a \in V$
6. $AV_{v'_e} := |N_e| / |M_e|$
7. $CAC_{v'_e} := (|N_e| / |M_e|) \cdot (|N_e| / |V'|)$
8. $CU_{v'_e} := (|N_e| / |V'|)^2 - (|M_e| / |V|)^2$
9. $AV_v := AV_v + AV_{v'_e}$
10. $CAC_v := CAC_v + CAC_{v'_e}$
11. $CU_v := CU_v + CU_{v'_e}$
12. **end for**
13. **end for**

Output: AV_v, CAC_v, CU_v for all $v \in \{T \cup C\}$

Homogeneity

Input: $DG = \langle V, E, P \rangle, e \in E$

1. **for all** $v \in \{C\}$ **do**
2. $V' :=$ the set of all $v' : v' \subseteq v$
3. **for all** $(v', v'') : v' \in V' \wedge v'' \in V'$ **do**
4. $V'_e := \{v'_e : \exists \langle v'_e, e, v' \rangle\}$
5. $V''_e := \{v''_e : \exists \langle v''_e, e, v'' \rangle\}$
6. $I := V'_e \cap V''_e$
7. $U := V'_e \cup V''_e$
8. $CN_{v',v''} := |I|$
9. $Jac_{v',v''} := |I| / |U|$
10. $Cos_{v',v''} := |I| / (\sqrt{|V'_e|} \cdot \sqrt{|V''_e|})$
11. $CN_v = CN_v + CN_{v',v''}$
12. $Jac_v = Jac_v + Jac_{v',v''}$
13. $Cos_v = Cos_v + Cos_{v',v''}$
14. **end for**
15. $CN_v = CN_v / (|V'| . (|V'| - 1) / 2)$
16. $Jac_v = Jac_v / (|V'| . (|V'| - 1) / 2)$
17. $Cos_v = Cos_v / (|V'| . (|V'| - 1) / 2)$
18. **end for**

Output: CN_v, Jac_v, Cos_v for all $v \in \{C\}$

Knowledge Anchor Derivation

- The above six metrics are calculated for each class entity in the graph, considering both its hierarchical and its domain-specific relationships.

- Hence, for each class we obtain a set of scores that rate that entity's suitability as a knowledge anchor.

- We select entities that have at least 50% non-zero scores, subject to the constraint that a knowledge anchor should have at least one non-zero score from the subset of hierarchical relationships and at least one non-zero score from the subset of domain-specific relationships.

Knowledge Anchor Derivation

- For example, the knowledge anchors identified within Occupation hierarchy of L4All2 were:

  Administrative_and_Secretarial_Occupations
  Administrative_Occupations
  Associate_Professional_and_Technical_Occupations
  Managers_and_Senior_Officials
  Personal_Service_Occupations
  Professional_Occupations
  Sales_and_Customer_Service_Occupations
  Teaching_and_Research_Professionals

4. Extending Flexible Querying with Knowledge Anchors

- Suppose the user is currently studying for a BSc in Information Systems and wishes to find out what possible future job choices there are by seeing what other people with qualifications in Information Systems, or similar, have gone on to do
- This can be undertaken by evaluating the following SPARQL^{AR} query
  - before running the query, the user elects (through a suitable GUI) to apply two of the edit operations that are available as part of the APPROX operator:
    - Insertion of an edge label; Substitution of an edge label;
  - and also selects one relaxation operation from those available as part of the RELAX operator:
    - Replacement of a subclass by its immediate superclass.

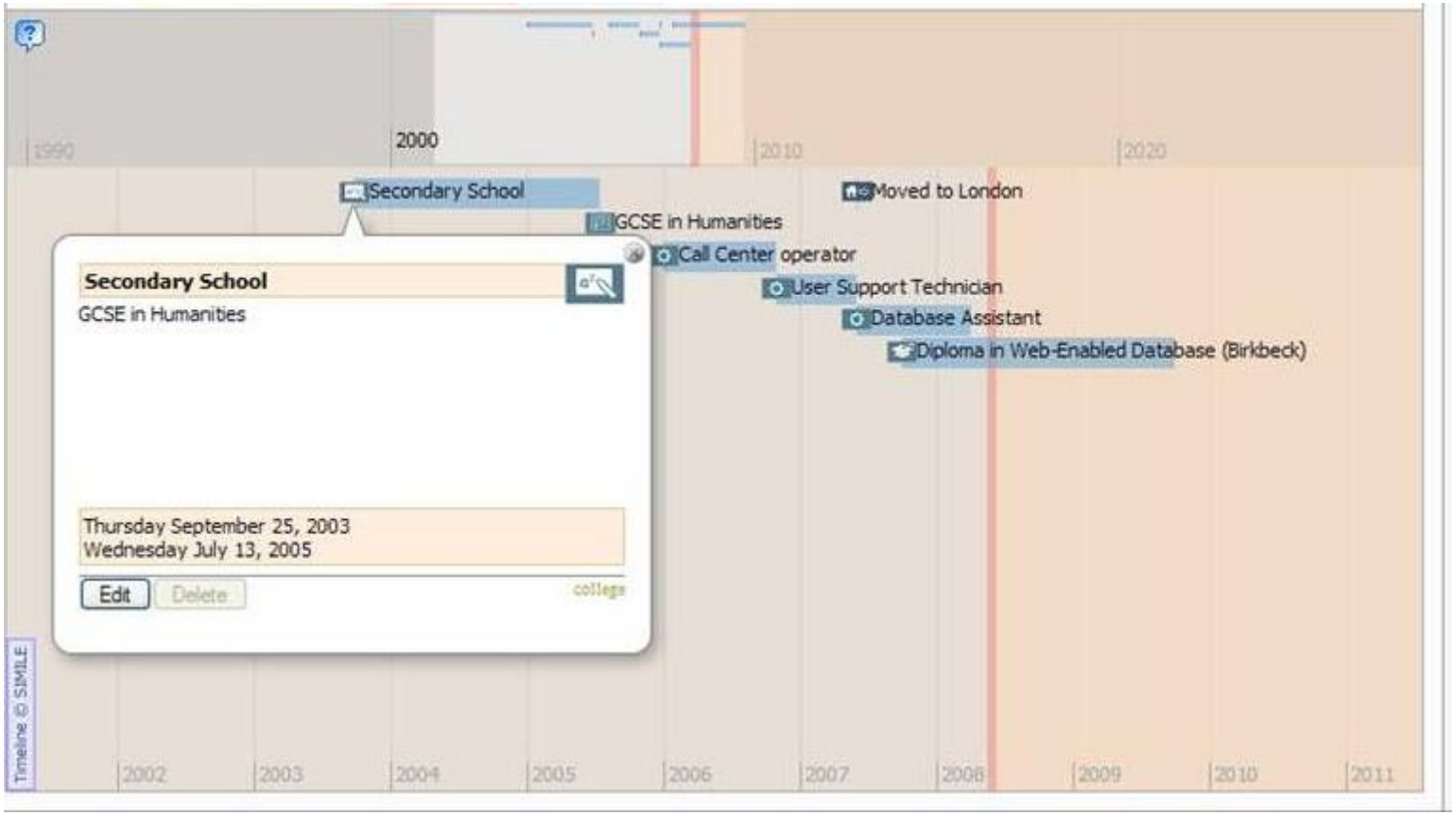
Extending Flexible Querying with Knowledge Anchors

```
SELECT ?WorkEp ?Occ
WHERE {?EdEp  rdf:type  <http://www.L4All.com/University_Episode>.
       RELAX ( ?EdEp <http://www.L4All.com/qualif>/rdf:type
                     <http://www.L4All.com/Information_Systems> ).
       APPROX( ?EdEp <http://www.L4All.com/prereq> ?WorkEp ).
       ?WorkEp  rdf:type  <http://www.L4All.com/Work_Episode>.
       ?WorkEp  <http://www.L4All.com/job>/rdf:type ?Occ}
```

First 20 exact answers

A_4_E_14_22	Research_Professionals
A_1_E_8_98	IT_User_Support_Technicians
A_3_E_5_37	Software_Professionals
A_7_E_7_92	Engineering_Technicians
A_7_E_7_4	Quality_Assurance_Technicians
A_7_E_7_60	Quality_Assurance_Technicians
A_2_E_6_11	Purchasing_Managers
A_8_E_5_88	Pensions_and_Insurance_Clerks
A_8_E_6_88	Physicists,_Geologists_and_Meterologists
A_1_E_8_58	IT_User_Support_Technicians
A_4_E_14_14	Registrars_and_Senior_Administrators_of_Educational_Establishments
A_7_E_7_52	Architectural_Technologists_and_Town_Planning_Technicians
A_4_E_14_78	Primary_and_Nursery_Education_Teaching_Professionals
A_7_E_7_20	Laboratory_Technicians
A_4_E_14_46	Scientific_Researchers
A_7_E_7_84	IT_Service_Delivery_Occupations
A_8_E_5_24	Stock_Control_Clerks
A_8_E_6_24	Electronics_Engineers
A_3_E_5_21	Software_Professionals
A_2_E_6_3	Advertising_and_Public_Relations_Managers

Clicking on an episode URI (in the first column of the results set) we can see a whole timeline



First 20 answers at distance 1

A_2_E_6_51	Personnel,_Training_and_Industrial_Relations_Managers	1
A_7_E_7_12	IT_User_Support_Technicians	1
A_8_E_5_56	Market_Research_Interviewers	1
A_7_E_7_44	Science_and_Engineering_Technicians	1
A_8_E_6_72	Civil_Engineers	1
A_8_E_5_64	Library_Assistants/Clerks	1
A_8_E_5_16	Transport_and_Distribution_Clerks	1
A_7_E_7_28	IT_Service_Delivery_Occupations	1
A_8_E_6_8	Mechanical_Engineers	1
A_8_E_5_96	Filing_and_Other_Records_Assistants/Clerks	1
A_2_E_6_35	Research_and_Development_Managers	1
A_8_E_6_40	Planning_and_Quality_Control_Engineers	1
A_4_E_14_6	Social_Science_Researchers	1
A_8_E_5_8	Library_Assistants/Clerks	1
A_1_E_8_50	IT_User_Support_Technicians	1
A_7_E_7_68	IT_User_Support_Technicians	1
A_1_E_8_74	IT_User_Support_Technicians	1
A_1_E_8_34	IT_User_Support_Technicians	1
A_7_E_7_60	Quality_Assurance_Technicians	1
A_3_E_5_69	Software_Professionals	1

First 20 answers at distance 2

A_2_E_6_51	Personnel,_Training_and_Industrial_Relations_Managers	2
A_7_E_7_12	IT_User_Support_Technicians	2
A_8_E_5_56	Market_Research_Interviewers	2
A_7_E_7_44	Science_and_Engineering_Technicians	2
A_8_E_6_72	Civil_Engineers	2
A_8_E_5_64	Library_Assistants/Clerks	2
A_8_E_5_16	Transport_and_Distribution_Clerks	2
A_7_E_7_28	IT_Service_Delivery_Occupations	2
A_8_E_6_8	Mechanical_Engineers	2
A_8_E_5_96	Filing_and_Other_Records_Assistants/Clerks	2
A_2_E_6_35	Research_and_Development_Managers	2
A_8_E_6_40	Planning_and_Quality_Control_Engineers	2
A_4_E_14_6	Social_Science_Researchers	2
A_8_E_5_8	Library_Assistants/Clerks	2
A_1_E_8_50	IT_User_Support_Technicians	2
A_7_E_7_68	IT_User_Support_Technicians	2
A_1_E_8_74	IT_User_Support_Technicians	2
A_1_E_8_34	IT_User_Support_Technicians	2
A_7_E_7_60	Quality_Assurance_Technicians	2
A_3_E_5_69	Software_Professionals	2

Extending Flexible Querying with Knowledge Anchors

- It is evident that, although it can return relevant and useful answers for the user, this kind of incremental flexible querying can easily result in information overload.
- Moreover, the user may be unfamiliar with some of the specialist terminology relating to occupations.
- The user will also gain little insight into the relationships between the different occupations being suggested and how they are categorised within the broader context of the Occupation hierarchy

- Repeating the above query and user interactions, we consider an alternative presentation of the results as *paths* within the Occupation hierachy, *rooted at the nearest Knowledge Anchor*

First 10 answers at dist. 0, under Knowledge Anchors

Associate_Professional_and_Technical_Occupations
  Science_and_Technology_Associate_Professional
    IT_Service_Delivery_Occupations
      IT_User_Support_Technicians A_1_E_8_98, A_1_E_8_58 (0)
  Science_and_Engineering_Technicians
    Engineering_Technicians A_7_E_7_92 (0)
    Quality_Assurance_Technicians A_7_E_7_4, A_7_E_7_60 (0)
Managers_and_Senior_Officials
  Corporate_Managers
    Functional_Managers
      Purchasing_Managers A_2_E_6_11 (0)
Professional_Occupations
  Science_and_Technology_Professionals
    Information_and_Communication_Technology_Professionals
      Software_Professionals A_3_E_5_37 (0)
    Science_Professionals
      Physicists,_Geologists_and_Meterologists A_8_E_6_88 (0)
Teaching_and_Research_Professionals
  Research_Professionals A_4_E_14_22 (0)
Administrative_Occupations
  Administrative_Occupations:_Records
    Pensions_and_Insurance_Clerks A_8_E_5_88 (0)

First 20 answers at dist. 0, under Knowledge Anchors

Associate_Professional_and_Technical_Occupations
  Science_and_Technology_Associate_Professional
    IT_Service_Delivery_Occupations **A_7_E_7_84 (0)**
      IT_User_Support_Technicians A_1_E_8_98, A_1_E_8_58 (0)
  Science_and_Engineering_Technicians
    Engineering_Technicians A_7_E_7_92 (0)
    Quality_Assurance_Technicians A_7_E_7_4, A_7_E_7_60 (0)
    Laboratory_Technicians A_7_E_7_20 (0)
  Draughtspersons_and_Building_Inspectors
    Architectural_Technologists_and_Town_Planning_Technicians A_7_E_7_52 (0)
Managers_and_Senior_Officials
  Corporate_Managers
    Functional_Managers
      Advertising_and_Public_Relations_Managers A_2_E_6_3 (0)
      Purchasing_Managers A_2_E_6_11 (0)
Professional_Occupations
  Science_and_Technology_Professionals
    Information_and_Communication_Technology_Professionals
      Software_Professionals A_3_E_5_37, **A_3_E_5_21** (0)
    Science_Professionals
      Physicists,_Geologists_and_Meterologists A_8_E_6_88 (0)
    Engineering_Professionals
      Electronics_Engineers A_8_E_6_24 (0)

Cont'd

Teaching_and_Research_Professionals
- Research_Professionals A_4_E_14_22 (0)
  - **Scientific_Researchers A_4_E_14_46 (0)**
- **Teaching_Professionals**
  - **Registrars_and_Senior_Administrators_of_Educational_Establishments A_4_E_14_14 (0)**
  - **Primary_and_Nursery_Education_Teaching_Professionals A_4_E_14_78 (0)**

Administrative_Occupations
- Administrative_Occupations:_Records
  - Pensions_and_Insurance_Clerks A_8_E_5_88 (0)
  - **Stock_Control_Clerks A_8_E_5_24 (0)**

First 20 answers at dist.0&1 under Knowledge Anchors

Associate_Professional_and_Technical_Occupations
  Science_and_Technology_Associate_Professional
    IT_Service_Delivery_Occupations A_7_E_7_84 (0); **A_7_E_7_28 (1)**
      IT_User_Support_Technicians A_1_E_8_98, A_1_E_8_58 (0);
        A_7_E_7_12, A_1_E_8_50, A_7_E_7_68, A_1_E_8_74, A_1_E_8_34 (1)
  Science_and_Engineering_Technicians **A_7_E_7_44 (1)**
    Engineering_Technicians A_7_E_7_92 (0)
    Quality_Assurance_Technicians A_7_E_7_4, A_7_E_7_60 (0); **A_7_E_7_60 (1)**
    Laboratory_Technicians A_7_E_7_20 (0)
  Draughtspersons_and_Building_Inspectors
    Architectural_Technologists_and_Town_Planning_Technicians A_7_E_7_52 (0)
Managers_and_Senior_Officials
  Corporate_Managers
    Functional_Managers
      Advertising_and_Public_Relations_Managers A_2_E_6_3 (0)
      Purchasing_Managers A_2_E_6_11 (0)
      Personnel,_Training_and_Industrial_Relations_Managers A_2_E_6_51 (1)
      Research_and_Development_Managers A_2_E_6_35 (1)
Professional_Occupations
  Science_and_Technology_Professionals
    Information_and_Communication_Technology_Professionals
      Software_Professionals A_3_E_5_37, A_3_E_5_21 (0); **A_3_E_5_69 (1) A_8_E_5_96 (1)**

Cont'd

- Science_Professionals
  - Physicists,_Geologists_and_Meterologists A_8_E_6_88 (0)
- Engineering_Professionals
  - Electronics_Engineers A_8_E_6_24 (0)
  - **Civil_Engineers A_8_E_6_72 (1)**
  - **Mechanical_Engineers A_8_E_6_8 (1)**
  - **Planning_and_Quality_Control_Engineers A_8_E_6_40 (1)**

Teaching_and_Research_Professionals

- Research_Professionals A_4_E_14_22 (0)
  - Scientific_Researchers A_4_E_14_46 (0)
  - **Social_Science_Researchers A_4_E_14_6 (1)**
- Teaching_Professionals
  - Registrars_and_Senior_Administrators_of_Educational_Establishments A_4_E_14_14 (0)
  - Primary_and_Nursery_Education_Teaching_Professionals A_4_E_14_78 (0)

Administrative_Occupations

- Administrative_Occupations:_Records
  - Pensions_and_Insurance_Clerks A_8_E_5_88 (0)
  - Stock_Control_Clerks A_8_E_5_24 (0)
  - **Market_Research_Interviewers A_8_E_5_56 (1)**
  - **Library_Assistants/Clerks A_8_E_5_64 (1)**
  - **Transport_and_Distribution_Clerks A_8_E_5_16 (1)**
  - **Filing_and_Other_Records_Assistants/Clerks A_8_E_5_96 (1)**

Extending Flexible Querying with Knowledge Anchors

- We see that the relationships between the occupations returned as query results are made explicit
- In parallel, the user can explore increasingly larger fragments of the Occupation hierarchy, each rooted at a Knowledge Anchor that may be more meaningful to the user than a specialist occupation.
  - We argue that this facilitates increasing awareness of possible relevant occupations by the user as compared with the purely linear presentation of results.
  - For future work we are planning trials with groups of students and practitioners from lifelong learning and careers guidance to investigate this hypothesis, through user evaluation activities that involve comparison of the two alternative forms of results presentation.

5. Conclusions and Future work

- This work addresses the challenge of supporting the *exploration of large knowledge graphs by users who are not experts in the domain*.
- We have proposed a *hybrid approach combining flexible graph querying and knowledge anchors*
  - flexible queries allow automatic expansion of query results by query approximation and query relaxation
  - knowledge anchors represent basic-level entities that are close to the user's cognitive structures; they are likely to be familiar to many users and can provide good starting points for introducing unfamiliar entities
- In our hybrid approach, we introduce knowledge anchors into query results by including *paths to the nearest knowledge anchor.*
- Our immediate plans are to develop interactive visualisations such as those illustrated earlier and to evaluate the approach with groups of students and practitioners
- We also plan to investigate other ways of hybridising flexible queries and knowledge anchors, e.g. for filtering or ranking query results.