Was hat sich bisher getan:

* Angefangen mit Proposal (Feedback von Exposee eingearbeitet)
* Angefangen mit Website (Unterschied zwischen Jars)
* Methoden fixiert (Design Science, Experiment, Expert-Interview, Literature Review)

Punkte zu besprechen:

* Success criteria bei Website: Schneller als shactor, usability mit fragen, wie oft man helfen muss. Ist das dann nicht quantitativ?
* Diff muss vorher schon gegeben sein, sonst nicht performant, kashif mail
  + Vorschlag: herausrechnen
* 3. rq weg: Einzige Idee: Interim resulsts: default shapes, die man nicht wieder konstruieren muss dann.

Nächste Schritte:

* Research Methods Paper schreiben, dazu bräuchte ich dann rq schon. Das ist vorbereitung für proposal im märz. Da möchte ich schon alles fixiert haben. Also ablauf, evaluierung, rq fix, methoden fix…
* Website weiterbauen
* Literatur Review kashif und evolve paper
* Nächstes Meeting 1. Oder 2. Jänner Woche

Notizen Meeting:

* RQ schicken vor Weihnachten zur Kontrolle
* Mail schreiben wegen VM
* Paper lesen
* Methoden definieren, qual quantitativ
  + Which shapes were added,
  + vergleich black box white box,
  + vergleich mit diff ohne diff
* Varianten RQ 3
  + How fast can we be if we don’t have the diff set before
    - Sollte rq 2 sein? Mit was?
  + Vergleich black box white box mit algo hacken
    - User zeit oder ausführungszeit? Mit sparql abfragen? Aber support und conf kann ich ja auch ausgeben. Aber das hätte ich bei anderen algorithmen nicht
  + Vergleich welche shapes warum weg oder dazugekommen sind
  + Given the shapes sets of two versions of a graphs, how can we efficiently use partial Sparql evaluation to explain the differences (diff noch ausgearbeitet) -> alt, ich habe eh qse daten

RQ1. What is an appropriate way to compare extracted shapes from  
different versions of a graph a user-friendly way?  
RQ2. How much faster is an adaption of the QSE-algorithm by initially  
calculating the difference between two graphs instead of running the  
algorithm twice for both versions? ->

RQ2 new: Given two versions of a graph (V1 and V2) and the changeset between these versions of, what is an appropriate way to adapt the QSE algorithm so that it calculates SHACL shapes of V2 by using V1 and the changeset?

* Wie viel schneller ist es, die Unterschiede zwischen den SHACL-Shapes herauszufinden, indem man zuerst den Algorithmus zweimal laufen lässt oder das Changeset verwendet?
  + Kein sinn, ist eh das erste schneller
  + Muss ich am anfagen machen? Oder kann ich das auch später bei den Maps machen? Anfang wahrscheinlich?
  + Muss ich mir noch überlegen
  + Ist das dann evaluation?
* Ist es schneller, zwei shacl shapes zu vergleichen (beliebiger algorithmus) und mit sparql abfragen die gründe zu finden oder die informationen aus dem QSE algorithmus zu verwenden?
* Wie können die gründe angegeben werden, warum shapes weggefallen/verändert/hinzugekommen sind

Schicken wenn schicken

Qual, quant, qual

Liefert das richtige ergebniss success criteria

Laufzeit vll später

Kashif fragen, wo ich das bekomme, mit daten. Statt sql abfrage

How fast can we be if we dont have the diff set before

Idee: einmal mit diff set, einmal ohne: wie lange dauert nur die berechnung ohne diff set, wenn man beides ausgeführt hat

1. Variatnte black box
2. 2. Algo hacken
3. Ergebnis: einmal hat man nur es für alle algorithm, einmal geht es schneller bei 2.
4. Man kann bewerten wie lange es dauert

2 wochen auf urlaub über weihnachten

50 gb – 100 gb

Kashif: wikidata ist groß

Email ncohmal schrebinn

Mo di mi nicht da

Nächste woche mit mail melden

8.1 10:00

Changesets sind oft da

Preprocessing

Sie hat noch paper: how does knowledge graphs evolve in knwoedge graphs (noch nicht öffentlich)

Metadaten sind auch da, weglassen

Schema ändert sich selten: selbst änderungen in testdaten einbauen

Anytime nicht

3. rq vereinfachen.

Given the shapes sets of two versions of a graphs, how can we efficiently use partial Sparql evaluation to explain the differences (diff noch ausgearbeitet)

Einzelne shpaes vergleichen zwishcen neuem und alten graph, auf grund dessen knan ich dnan sagen warum was weggefallen ist, verändert ist oder hinzugekomme ist. Enweder auf altem oder auf neuem machen je nach grund.

Auch mit experiment. Nur messen. Qualitativ -> ist das ergebnis richtig, wie lange hat es gedauert.

User study und qualitativ

Shactor passt

Vm und paper

Speicherplatz gedanken machen, 1 prozessor, wikipedia dump ansehen mal 10 + extra platz, ram 256, maximum nehmen (so viel nicht), linux

Alte Notizen:

3. rq weg: Interim resulsts: default shapes, die man nicht wieder konstruieren muss dann.

Qualitative evaluation: reicht expert interview oder muss ich auch fragebogen machen, um die usablity zu messen?

Success criteria?

Passt state of the art für sie? -> literature review später

Schneller, intuitiver, Vll zeit messen im vergleich zu dem ich es mit shactor mache?

Literature review machen

Success criteria: duration, wie oft man helfen muss, satisfaction

Ich bereite shactor vor mit den datensets und lasse sie dann mit einem einfachen beispiel die unterschiede finden, mit und ohne webapp, am schluss fragen, wie sie es fanden. Von vorne mit shapes generieren? Oder mit fertigen shapes? Vll mehr fokus auf mehrere vergleiche

Rq2: wie relevant, kashif sagt, dass sich schema selten ändert

Kathi fragen welche methode es ist, wenn man sachen messen muss, qualitativ aber doch usablilty

Ist dann die methode qualitativ und quantitativ?

Alte Notizen:

Difference hat zwei listen: triples hinzugefügt und gelöscht

Diese werden dann vom ursprünglichen graphen hinzugefügt oder gelöscht. Dann wird qse ganz normal ausgeführt. Laufzeit:

Laufzeit normal: O(2n) + O (2m)

Laufzeit-Differenz: O(2\*n\*m) + O(2n) (??) weiß nicht wie lange difference von jena braucht

Mit sets:

Zwei mal converten O(n)+O(m), Zwei mal remove all O(n)+O(m) + durchlaufen O(2n)

Kann nicht schneller sein, außer es gibt eine Variante, die Change sets schneller zu bekommen als 2m

Könnte schneller sein, wenn man changeset hinauflädt (added triples und removed triples)

Könnte auch das changeset nehmen und dann direkt in den algorithmus eingreifen mit den daten

Allgemein nachfragen

Allgemein nachfragen: ist input immer zwei graphen oder auch nur changeset? Ist es dann rdfpatch oder einfach 2 versionen?

Interim resulsts: default shapes, die man nicht wieder konstruieren muss dann.

Shactor mit film.nt:

Wenn Datenset ausgeführt wird, werden zuerst mittels type die klassen herausgefiltert.

* Parser, entityExtraction
  + Befüllt entitydatahashmap: Map<Node, EntityData>
  + Befüllt Map<Integer, Integer> classEntityCount
  + (Schritt 1)
* Nach Klassenauswahl (geht graph zum zweiten mal durch)
  + Parser, entityConstraintsExtraction
  + Befüllt entityDataHashMap weiter (Schritt 2)
  + Befüllt classToPropWithObjTypes
* extractSHACLShapes
* ShapesExtractor: constructDefaultShapes

Welche Results kann ich verwenden, außer die shapes, die schon konstruiert worden sind? Kashif fragen?

Wie funktionieren Changesets für große Graphen? Ist es dann nicht komplett unperformant, zuerst die genaue Differenz zu berechnen, wenn später eh gesamplet wird?

Exakt und approx nochmal besprechen

Kashif fragen

Qualitative evaluation: reicht expert interview oder muss ich auch fragebogen machen, um die usablity zu messen?

Fragen:

<https://afs.github.io/rdf-delta/cmds.html>

Passt state of the art für sie?

Vorher nachher vergleich für usability?

Huemer fragen

Vll goals kürzen

Vale schreiben, endi,

passen success criteria?

Kathi paper schicken

Rq1 passt nicht, weil how to question. Sollte ich dann davor studie machen und danach und dann qualitativ vergleichen?

Rq3 passt nicht, weil nicht messbar mit methode

Kann ich auch kein Design Science nehmen, ich muss ja sowieso alles evaluieren.

Proposal Review Problem Statement: ♣ What is the problem that does not exist anymore once the thesis is (successfully) finished? Try to describe it in one or two sentences. ♣ Does the proposal specify any measurable characteristics of the current situation that need to be solved? Are metrics (KPIs) for these characteristics defined? Are there even values for these characteristics given? If so, outline them.

Goals and Expected Outcome: ♣ What is the proposed solution (to overcome the problem)? Try to describe it in one or two sentences. ♣ Are the characteristics of the solution well defined? Are metrics (KPIs) for these characteristics specified? If so. Outline them. ♣ When is the solution considered a success?

Research Questions: ♣ Is it possible to provide scientific evidence for all answers to the research questions? ♣ Are there any research questions that are not precise (Yes/No questions, “How to develop sth”, etc.)? ♣ If the goal is to develop something appropriate/suitable/etc., does it become clear what appropriate/suitable/etc. means?

Research Methods: ♣ Are the envisioned research method(s) appropriate to reach the expected research results? ♣ Are the methodological steps to be followed clear? If so, outline them. ♣ Is each methodological step based on a research method known from the literature? If so, mention the corresponding paper from the literature describing the underlying research method for each methodological step. ♣ [Only for a thesis that follows Design Science: Design Science requires by definition the application of rigorous research methods. Is it clear which research methods are used within the Design Science framework?]

Evaluation: ♣ Which research method is used for the evaluation? ♣ Are the metrics to be used in the evaluation clear? Which are they? ♣ Is the evaluation appropriate? Does the evaluation provide scientific evidence by coming up with results based on KPIs that demonstrate that the solution solves the identified problem? Provide arguments. [Note, an approach could also fail - so after the thesis, the instantiation of the proposed evaluation may show that the problem is still unsolved.]

State of the Art: ♣ Does the elaboration on state of the art show that the student knows the relevant literature in the field? Provide arguments. ♣ Is there any literature missing?

Relevance to the Curriculum: ♣ Do you consider the thesis relevant for the given curriculum? Provide arguments. ♣ Is the relevance argued by a somewhat arbitrary list of courses (without arguing why each course is relevant), or does the section elaborate on the knowledge necessary to conduct the thesis and, then, show in which courses this know-how is taught?

Overall review ♣ Do you consider the proposal concise? ♣ Is the proposal of excessive length and, thus, should be shortened? First, note a proposal should be limited to about five pages (9000 characters including whitespaces) without the literature list. ♣ Is the style of writing appropriate? Is there a clear line of arguments? ♣ What are the strengths of the proposal? ♣ What are the limitations of the proposal? ♣ What should be revised?

Seaborn und rdfpatch ist eher für changesets, wenn man selber changes durchführt.

Apache jena difference compared die zwei graphen ganz einfach.

Todo: kashif mail besprechen

Proposal fragen überarbeiten

Andere proposals lesen

Alles überarbeiten

3. rq weg?

Literature review machen

Success criteria: duration, wie oft man helfen muss, satisfaction

Ich bereite shactor vor mit den datensets und lasse sie dann mit einem einfachen beispiel die unterschiede finden, mit und ohne webapp, am schluss fragen, wie sie es fanden. Von vorne mit shapes generieren? Oder mit fertigen shapes? Vll mehr fokus auf mehrere vergleiche

Rq2: wie relevant, kashif sagt, dass sich schema selten ändert

Kathi fragen welche methode es ist, wenn man sachen messen muss, qualitativ aber doch usablilty

Ist dann die methode qualitativ und quantitativ?