How to write a good state of the art: should it be the first step of your thesis?

Cassia Trojahn dos Santos

Institut de Recherche en Informatique de Toulouse, France
cassia.trojahn@irit.fr

Meeting of Young Researchers in Artificial Intelligence (RJCIA), June (2019)

Outline

Outline

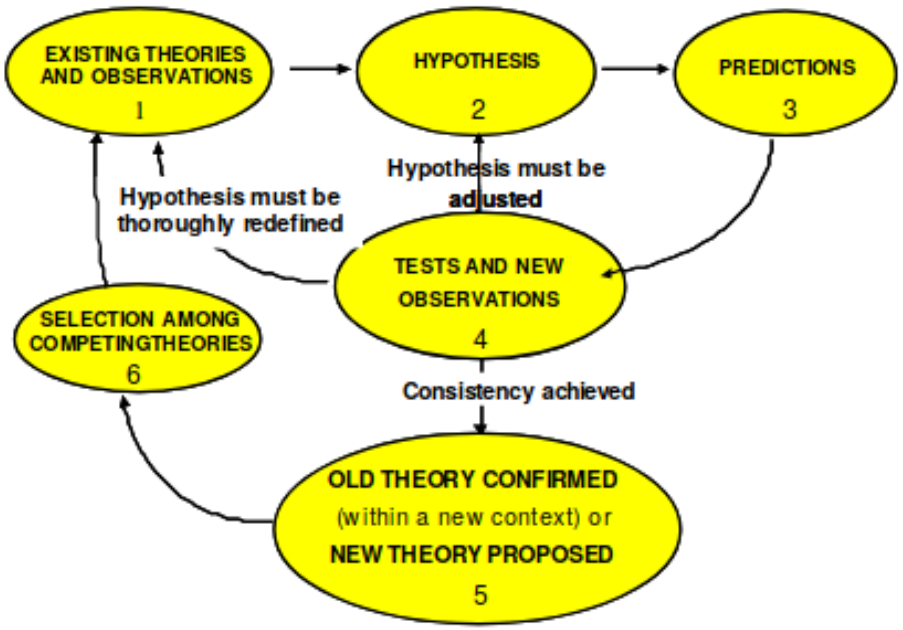
PhD

- A PhD is an (individual) research project involving advanced scholarship, that makes an original **contribution to knowledge**
- PhD: doctor of philosophy
  - Doctor: from the classical Latin "Teacher" (to show, teach, cause to know)
  - Philosophiae: from the Greek, meaning "love of knowledge", "pursuit of wisdom", "systematic investigation"

Source: V. Tamma (EKAW DC 2018)

The Scientific Method

- "The scientific method is the logical scheme used by scientists searching for answers to the questions posed within science"

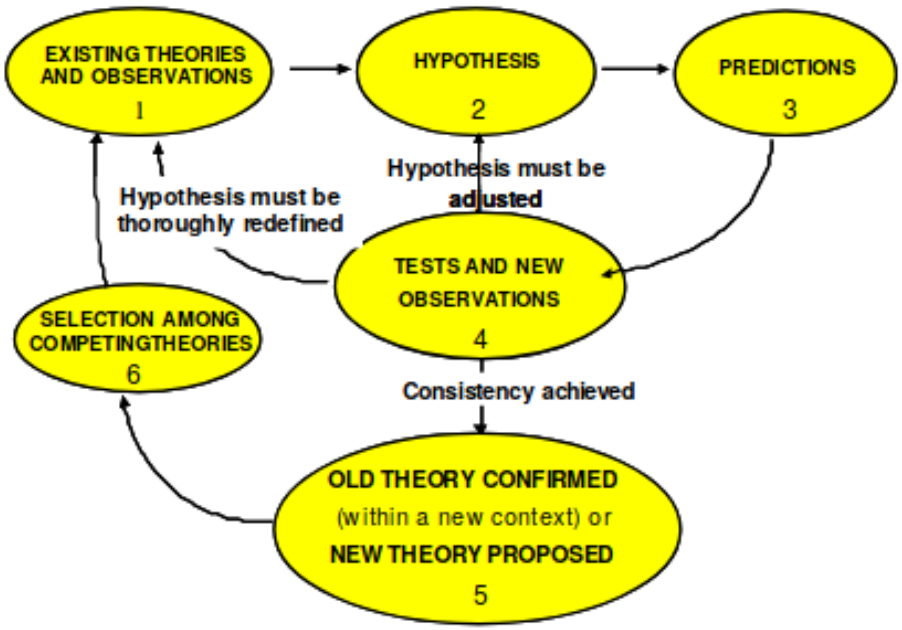


THE SCIENTIFIC METHOD

Source: Gordana DODIG-CRNKOVIC. Scientific Methods in Computer Science. (2002)

The Scientific Method

- "The scientific method is the logical scheme used by scientists searching for answers to the questions posed within science"



THE SCIENTIFIC METHOD

First step: Pose the question in the context of **existing knowledge** (theory & observations).

Gordana DODIG-CRNKOVIC. Scientific Methods in Computer Science. (2002)

The brief

Talk about How to write a good state of the art: should be it the first step of your thesis ?

Outline

State of the art/Literature review

"Systematic literature review on the state of the art and future research work in anonymous communications systems"

(M.Nia and A. Martnez, Computers & Electrical Engineering, Vol. 69, July 2018, Elsevier)

State of the art/Literature review

"Systematic literature review on the state of the art and future research work in anonymous communications systems"

(M.Nia and A. Martnez, Computers & Electrical Engineering, Vol. 69, July 2018, Elsevier)

- Literature review:
  - everything that is relevant

    (Inspired from D. Ridley. The Literature Review. Sage Study Skills, 2008)

- State of the art:
  - the highest degree of development of an art or technique at a particular time

    (Source: WordNet)

State of the art/Literature review

- "The 'literature review' is the part of the thesis where there is extensive reference to related research and theory in your field; it is where connections are made between the source texts that you draw on and you position your research among these sources..."
- "The 'literature review' is where you identify the theories and previous research which have includes in your choice of research topic ... you can use the literature to support your identification of a problem to research and to illustrate that there is a gap in previous research which needs to be filled ..."

D. Ridley. The Literature Review: A Step-by-step Guide for Students. Sage Study Skills, 2nd edition (2008)

State of the art/Literature review

- Describes the **knowledge** about the studied **matter** through the **analysis** of similar or related **published work**
- Provides a **comprehensive overview** of what was done, what has been done in the field and what should be further investigated

State of the art/Literature review

- Describes the **knowledge** about the studied **matter** through the **analysis** of similar or related **published work**
- Provides a **comprehensive overview** of what was done, what has been done in the field and what should be further investigated

$\Rightarrow$ Helps formulating the **problems and hypothesis the thesis** intends to address !

State of the art/Literature review

- Describes the **knowledge** about the studied **matter** through the **analysis** of similar or related **published work**
- Provides a **comprehensive overview** of what was done, what has been done in the field and what should be further investigated

$\Rightarrow$ Helps formulating the **problems and hypothesis the thesis** intends to address !

$\Rightarrow$ Producing a (*good*) state of the art might be considered the main initial step of a PhD thesis !

State of the art/Literature review

- "The literature review also refers to the process involving in creating the review that appears in your dissertation or thesis ..."

  D. Ridley. The Literature Review. Sage Study Skills, 2nd edition (2008)

- It is a **process** and an **product** !

Outline

The main steps of the whole process

1. Define the scope
2. Identify pertinent sources
3. Collect information (read, organise, take notes, keep track of citations)
4. Write, revise
5. Back to 2 or 3

(Source: Pixnio)

Be ready !

- Process that involves analysing, comparing, evaluating, criticising, discovering relationships, and classifying existing works ... and identifying gaps and contributions !

  $\Rightarrow$ many (MANY!) hours of reading and content organisation

It is not a linear process !

- Involves many interactions ($\Rightarrow$ reading a new paper, for example, allows for having new ideas on how organising the content and hence reviewing things)
- Requires to be continuously and incrementally refreshed (state of the art)
- State of the art finishes when the thesis finishes? **Almost**!

Some advice

- Make sure you include all relevant sources and those recent from major conferences in the field (ex., IJCAI, ECAI, AAAI, etc.)
- Check other thesis on the topic (oatd.org, openthesis.org, etc)
- Identify communities on the topic (specific workshops on national ad international conferences)
- Choose scholarly articles (and pay attention to the impact factor of journals, for instance)
- Create alerts, follow researchers, keep informed !

Some advice

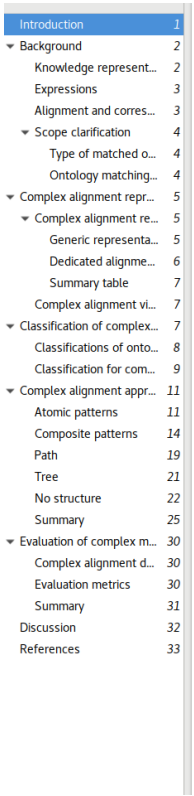
- Take notes (together with bibliographic entries, keep track of the pages you cite ... at reading time!)
- Use dedicated tools for managing your references (Zotero, Mendeley, BibDesk, Citavi, etc...)
- Normalise bib entries, group them, share them across projects

Outline

Different types

1. Stand-alone paper (surveys)
2. Part of a conference or journal paper
3. Part of your thesis

Different types



Semantic Web 0 (0) 1
IOS Press

Survey on complex ontology matching

Elodie Thiéblin *, Ollivier Haemmerlé, Nathalie Hernandez, Cassia Trojahn
Equipe MELODI, Institut de Recherche en Informatique de Toulouse, Toulouse, France
E-mails: elodie.thieblin@irit.fr, ollivier.haemmerle@irit.fr, nathalie.hernandez@irit.fr, cassia.trojahn@irit.fr

Editor: Marta Sabou, Technical University of Vienna, Austria
Solicited reviews: Antoine Zimmermann, Mines Saint-Etienne, Université Jean Monnet, IOGS, CNRS, UMR 5516 lHC, Institut Henri Fayol, France; Catia Pesquita, LASIGE, Faculdade de Ciências da Universidade de Lisboa, Portugal; anonymous reviewer

Abstract. Simple ontology alignments, largely studied in the literature, link a single entity of a source ontology to a single entity of a target ontology. A limitation of these alignments is their lack of expressiveness which can be overcome by complex alignments. While diverse state-of-the-art surveys mainly review the matching approaches in general, to the best of our knowledge, there is no study of the specificities of the complex matching problem. In this paper, a review of the different complex matching approaches is provided. It proposes a classification of the complex matching approaches based on their specificities (i.e., type of correspondences, guiding structure). The evaluation aspects and the limitations of these approaches are also discussed. Insights for future work in the field are provided.

Keywords: ontology matching, complex alignment, survey, schema matching

`http://www.semantic-web-journal.net/content/survey-complex-ontology-matching`

Different types

1 Introduction

Ontology matching is an essential task for the management of the semantic heterogeneity in open environments. The matching process aims at generating a set of correspondences (i.e., an alignment) between the entities of different ontologies. Two 'paradigms' organise the field. While approaches generating simple correspondences are limited in expressiveness by linking single entities, complex matching approaches are able to generate correspondences which better express the relationships between entities of different ontologies. Earlier works have introduced the need for complex alignments [15,34]. Different approaches for generating such complex alignments have been proposed in the literature. While the proposal of [23,24] relies on correspondence patterns, the one in [13] uses knowledge-rules in Markov-Logic Networks. Those in [20,21,35] rely on statistical methods and correspondence patterns and the one in [18] deals with genetic programming. Finally, the approach in [22] uses path-finding algorithms combined with statistical techniques. Despite the progress in the field, there is a lack of reference alignment sets on which the complex approaches can be systematically evaluated. Most efforts on evaluation are still dedicated to the matching approaches dealing with simple alignments. Systematic evaluation of them has been carried out over the last fifteen years in the context of the Ontology Align-

E. Thiéblin, O. Haemmerlé, N. Hernandez, C. Trojahn: Task-Oriented Complex Ontology Alignment: Two Alignment Evaluation Sets. ESWC 2018: 655-670

Different types: part of a paper

to find correspondences with transformation functions between two knowledge bases. The one in [22] uses a path-finding algorithm to find correspondences between two knowledge bases with common instances. The correspondences found by this approach are of the form *property path* = *property path*.

The different approaches discussed above are generic in the sense that they generate alignments regardless to a specific task.

3.2 Evaluation of matchers

Alignments generated by matchers can be evaluated in different ways [6]. One way consists in comparing alignments to reference ones (gold standard). However, constructing such references is a time-consuming task. In the lack of such resources, alternatives include manual labelling on sample alignments, alignment coherence measurements [17] and checking the conservativity principle violation of alignments [27]. Furthermore, the quality of a matcher can be assessed regarding its suitability for a specific task or application [12,10,28]. Finally, alternative approaches for validating alignments consider the generation of natural language questions to support end-users in the validation task [1] or validation of correspondences using graph-based algorithms in a semi-automatic way [26].

While matching evaluation has been focused on simple alignments, complex evaluation has been addressed to a lesser extent. Although a large spectrum of matching cases are proposed in the OAEI, e.g., involving synthetically generated or real case datasets with large or domain-specific ontologies, these datasets are limited to alignments with simple correspondences.

E. Thiéblin, O. Haemmerlé, N. Hernandez, C. Trojahn:Task-Oriented Complex Ontology Alignment: Two Alignment Evaluation Sets. ESWC 2018: 655-670

Different types: chapter of your thesis

Table 4.2: Approach positioning with regards to Tables 3.4, 3.5, 3.6 and 3.7

Approach	Type of Knowledge Representation Model	Additional Input
CANARD	OWL ontology to OWL ontology	CQAs

Approach	(s:c)	(c:s)	(c:c)	Logic	Transfo	Block	Correspondence format
CANARD	•	•	•	•			EDOAL, DL

Approach	Guiding struct
CANARD	No structure, Path

Approach	Formal resour	Inform
CANARD	•	

Benchmark	Type of evaluation	Reference	Anchoring	Comparison	Scoring
OAEI simple tracks	Automatic	Alignment	IRI	Syntactic	Classical
OA4QA [Solimando 2014b]	Automatic	Alignment	Source Query	Instance-based	Comp. val.
[Hollink 2008]	Manual	Query	Source Query	Instance-based	Classical/ Comp. val.
[Walshe 2016]	Manual	Alignment	Source IRI	Syntactic	Classical
[Parundekar 2012]	Manual	Alignment	Source IRI	Syntactic	Classical
[Thiéblin 2018]	Manual	Alignment	Source IRI	Semantic	Classical
Complex conf. [Thiéblin 2018a]	Manual	Alignment	Source IRI	Semantic	Classical
GeoLink [Thiéblin 2018a]	Manual	Alignment	Source IRI	Syntactic	Classical
Hydrography [Thiéblin 2018a]	Manual	Alignment	Source IRI	Syntactic	Classical
Taxon [Thiéblin 2018a]	Manual	Query	Source Query	Semantic	Classical
Populated conf.	Automatic	Query	Source Query	Instance-based	Many

Different types: chapter of your thesis

- The survey paper that you wrote in the first year of your thesis :-)
- You you thank your advisor for that ! No energy for doing the whole state of the art at the end ... believe me !

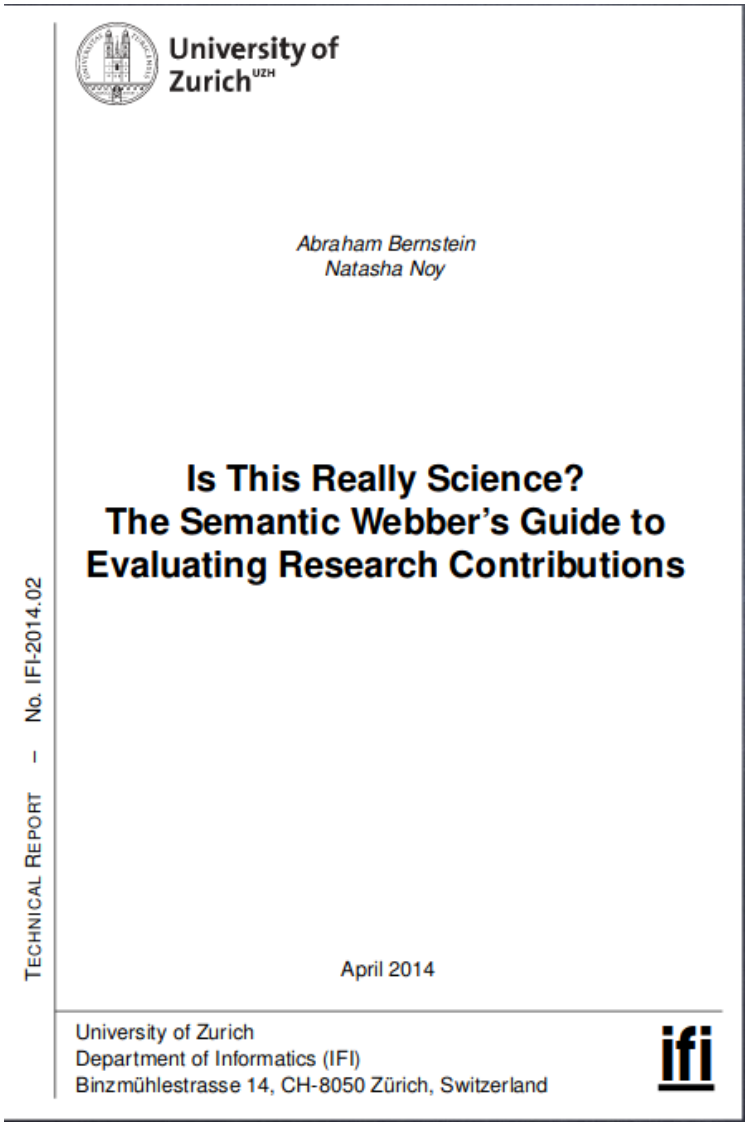
To conclude with some advice ...

- Read, read, read... and read
- Read anything that captures your imagination
- Read with questions in mind:
  - "How can I use this"
  - "Does this really do what the authors claim?"
  - "Do I understand the results in the paper?"
- Talk about your research
  - To your supervisor(s), to your colleagues, to students in other departments

(Inspired from V. Tamma (EKAW DC 2018)

- Exchange with your supervisor (s) !
- Start you thesis with a state of the art (and literature review) !

Further reading ...



University of Zurich[UZH]

Abraham Bernstein
Natasha Noy

Is This Really Science?
The Semantic Webber's Guide to Evaluating Research Contributions

TECHNICAL REPORT – No. IFI-2014.02

April 2014

University of Zurich
Department of Informatics (IFI)
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

ifi

http://iswc2018.semanticweb.org/wp-content/uploads/2018/10/2018-ISWC-DC.pdf