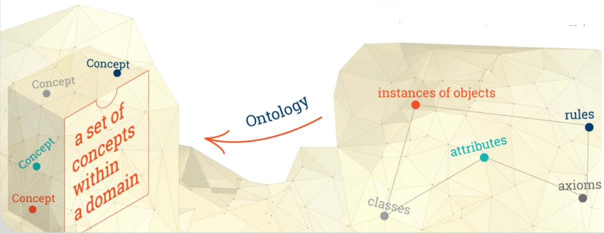
Proseminar work:
Investigation of Ontologies in Software-Engineering-Meta-Research

Advisor: Dipl.-Inform. Angelika Kaplan

Dmitrii Seletkov | July 2, 2020

SOFTWARE DESIGN AND QUALITY, INSTITUTE FOR PROGRAM STRUCTURES AND DATA ORGANIZATION, KIT DEPARTMENT OF INFORMATICS



Outline

1. Motivation
2. Foundations
   - Ontologies in Computer Science
   - Meta-Research
3. Ontologies in Software-Engineering-Meta-Research
4. Ontology based systems in scientific search
5. Conclusion

Motivation

- Retrieving and transferring Knowledge: essential part of human being

Motivation

- Retrieving and transferring Knowledge: essential part of human being
- **But**: the most amount of Knowledge is understandable only for humans

Motivation

Karlsruhe Institute of Technology

- Retrieving and transferring Knowledge: essential part of human being
- **But**: the most amount of Knowledge is understandable only for humans
- Ontologies make Knowledge understandable for computers as well, that provides:
  - Supporting humans in Knowledge transferring process
  - Opportunity to analyze and generate new knowledge automatically by machines

Motivation

- Retrieving and transferring Knowledge: essential part of human being
- **But**: the most amount of Knowledge is understandable only for humans
- Ontologies make Knowledge understandable for computers as well, that provides:
  - Supporting humans in Knowledge transferring process
  - Opportunity to analyze and generate new knowledge automatically by machines
- Useful for Software Engineering
  - Encapsulate the results of thousands Software Engineering experiments
  - Make possible to analyze them and find out the best Software Engineering practice

Ontology in Computer Science

Requirements for Knowledge

- Knowledge must be:
  - Adequate and representative
  - Machine readable and able to elaborate new Knowledge
- **But:** Information $\neq$ Knowledge

Ontology in Computer Science

Requirements for Knowledge

- Knowledge must be:
  - Adequate and representative
  - Machine readable and able to elaborate new Knowledge
- **But:** Information $\neq$ Knowledge

Def. Ontology in Computer Science

- "an explicit specification of a conceptualization"
- conceptualization: abstract model of some knowledge domain
- explicit specification: classes, concepts, terms

Ontology languages

Requirements for Ontology languages

- clear, computer readable syntax with formal semantics (unique meaning)
- effectively computer analyzable and sufficient descriptive power

Ontology languages

Karlsruhe Institute of Technology

Requirements for Ontology languages

- clear, computer readable syntax with formal semantics (unique meaning)
- effectively computer analyzable and sufficient descriptive power

Well-known examples

- ER-Diagrams and UML-Diagrams
- good for understanding and representing of Knowledge, but still made for humans

Ontology languages

Karlsruhe Institute of Technology

Requirements for Ontology languages

- clear, computer readable syntax with formal semantics (unique meaning)
- effectively computer analyzable and sufficient descriptive power

Well-known examples

- ER-Diagrams and UML-Diagrams
- good for understanding and representing of Knowledge, but still made for humans

Resource Description Framework (RDF)

- Natural and logical XML-based syntax
- Standard by W3C [citate]

RDF (1)

Principles of RDF

- Everything is representable through triples *<subject, predicate, object>*
- *subject, predicate* and *object* are resources identified by a unique reference (URI)
- Every RDF-Statement consists of resources

RDF (1)

Principles of RDF

- Everything is representable through triples *<subject, predicate, object>*
- *subject, predicate* and *object* are resources identified by a unique reference (URI)
- Every RDF-Statement consists of resources

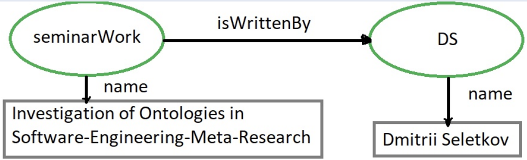
RDF example: the current proseminar work

- *<seminarWork, name, Investigation of Ontologies in Software-Engineering-Meta-Research>*
- *<seminarWork, isWrittenBy, DS>*
- *<DS, name, Dmitrii Seletkov>*

RDF (2)

RDF example: the current proseminar work

<seminarWork, name, Investigation of Ontologies in Software-Engineering-Meta-Research>
<seminarWork, isWrittenBy, DS>
<DS, name, Dmitrii Seletkov>




```
<rdf:Description rdf:about="#seminarWork">
    <csclub:isWrittenBy rdf:resource="#DS"/>
    <csclub:name>Investigation of Ontologies in Software-Engineering-Meta-Research</csclub:name>
</rdf:Description>
<rdf:Description rdf:about="#DS">
    <csclub:name>Dmitrii Seletkov</csclub:name>
</rdf:Description>
```

RDF (3)

Restrictions of RDF

- No local restrictions on a knowledge domain e.g. predicate *hasChild* applied on a class *Person* or a class *Animal* (No semantic difference)
- No transitive, symmetric or reverse predicates e.g. if A *touches* B, then B *touches* A
- Many others

RDF (3)

Restrictions of RDF

- No local restrictions on a knowledge domain e.g. predicate *hasChild* applied on a class *Person* or a class *Animal* (No semantic difference)
- No transitive, symmetric or reverse predicates e.g. if A *touches* B, then B *touches* A
- Many others

Description Logic (DL)

- Family of knowledge representation languages, motivated by restrictions of RDF
- Has formal semantics and instruments of logical analysis
- Has different dialects and implementations such as OWL
- OWL: Ontology Web Language, current standard for Semantic Web

Description Logic (1)

DL vs. RDF

Clear separation between Terminology (concepts and relations in ontology schema) and Assertions (concrete individuals that pertain to the concepts in the schema).

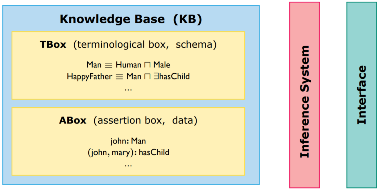


Figure: Architecture of Description Logic Konev 2010, lecture 8, slide 1

Description Logic (2)

EL DL

- Names of *Concepts* (also classes) such as *Person, Female, ...*
- Names of *roles* (also relations) such as *isWrittenBy, hasChild, ...*
- Concept *Thing* $\top$, Everything (every concept) is a *Thing*
- Conjunction (logical and) $\sqcap$; Existence quantor $\exists$
- Definition or equivalence $\equiv$; Primitive definition or subset $\sqsubseteq$

Description Logic (2)

Karlsruhe Institute of Technology

EL DL

- Names of *Concepts* (also classes) such as *Person, Female, ...*
- Names of *roles* (also relations) such as *isWrittenBy, hasChild, ...*
- Concept *Thing* $\top$, Everything (every concept) is a *Thing*
- Conjunction (logical and) $\sqcap$; Existence quantor $\exists$
- Definition or equivalence $\equiv$; Primitive definition or subset $\sqsubseteq$

Examples

- $Person \sqcap Female$

Description Logic (2)

EL DL

- Names of *Concepts* (also classes) such as *Person, Female, ...*
- Names of *roles* (also relations) such as *isWrittenBy, hasChild, ...*
- Concept *Thing* $\top$, Everything (every concept) is a *Thing*
- Conjunction (logical and) $\sqcap$; Existence quantor $\exists$
- Definition or equivalence $\equiv$; Primitive definition or subset $\sqsubseteq$

Examples

- $Person \sqcap Female$ (= woman)
- $Person \sqcap \exists hasChild.\top$

Description Logic (2)

EL DL

- Names of *Concepts* (also classes) such as *Person, Female, ...*
- Names of *roles* (also relations) such as *isWrittenBy, hasChild, ...*
- Concept *Thing* $\top$, Everything (every concept) is a *Thing*
- Conjunction (logical and) $\sqcap$; Existence quantor $\exists$
- Definition or equivalence $\equiv$; Primitive definition or subset $\sqsubseteq$

Examples

- $Person \sqcap Female$ (= woman)
- $Person \sqcap \exists hasChild.\top$ (= parent)
- $Person \sqcap \exists hasChild.\top \sqcap \exists hasChild.\top$

Description Logic (2)

EL DL

- Names of *Concepts* (also classes) such as *Person, Female, ...*
- Names of *roles* (also relations) such as *isWrittenBy, hasChild, ...*
- Concept *Thing* $\top$, Everything (every concept) is a *Thing*
- Conjunction (logical and) $\sqcap$; Existence quantor $\exists$
- Definition or equivalence $\equiv$; Primitive definition or subset $\sqsubseteq$

Examples

- $Person \sqcap Female$ (= woman)
- $Person \sqcap \exists hasChild.\top$ (= parent)
- $Person \sqcap \exists hasChild.\top \sqcap \exists hasChild.\top$ (= grandparent)

Ontology Web Language (OWL)

- OWL2: the current Ontology language standard
- Based on DL concepts and XML
- Fully compatible for already existing RDF-based ontology

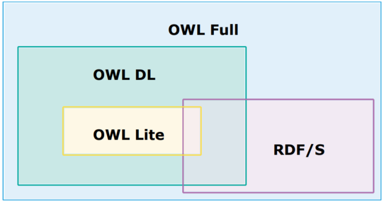


Figure: OWL dialects Konev 2010, lecture 8, slide 40

Meta-research (1)

Motivation

- Research practices suffer from lack of systematization and inefficiency
- Problems with data sharing, replications of experiments and their ownership
- Urgent need of the science for the evaluation of diverse researches to improve the existing research practices and create the new ones

Meta-research (1)

Motivation

- Research practices suffer from lack of systematization and inefficiency
- Problems with data sharing, replications of experiments and their ownership
- Urgent need of the science for the evaluation of diverse researches to improve the existing research practices and create the new ones

Def. Meta-Research

"The use of scientific methodology to study science itself"

Meta-Research (2)

Areas of Meta-Research (by Ioannidis et al. n.d.)

1. **Methods**: practices for performing research (**e.g.** study design, methods, statistics).
2. **Reporting**: publications of standards and study registrations (**e.g.** study registration, information to patients, public and policy-makers)
3. **Reproducibility:** methods for verifying research (**e.g.** sharing data and methods, replicability)
4. **Evaluation**: approvements for scientific quality (**e.g.** pre- and post-publication peer reviews, research funding criteria).
5. **Incentives**: rewards and penalties for research (**e.g.** promotion criteria, penalties in research evaluation).

Ontologies in Software-Engineering-Meta-Research

Classification of empirical studies (by Almeida Biolchini et al. 2007)

- **Primary study**: evaluate the hypothesis formulated by the researcher, to be tested under well-established conditions. Includes **Controlled Experiments**, survey and case studies
- **Secondary study**: comparisons between individual investigations of primary studies to generalize results. Includes **Systematic Reviews**

Ontologies in Software-Engineering-Meta-Research

Classification of empirical studies (by Almeida Biolchini et al. 2007)

- **Primary study**: evaluate the hypothesis formulated by the researcher, to be tested under well-established conditions. Includes **Controlled Experiments**, survey and case studies
- **Secondary study**: comparisons between individual investigations of primary studies to generalize results. Includes **Systematic Reviews**

Evidence-based Software Engineering (EBSE)

- Originates from Evidence-based Medicine
- Purpose: determine what SE practice works, when, where and which tools and standards needed
- The main instrument: Systematic Reviews (SRs)

Ontology to support systematic reviews in Software Engineering

Problem

- Major challenge to strengthen the foundations of SE:
  produce knowledge that can be based on scientific methodology

Ontology to support systematic reviews in Software Engineering

Problem

- Major challenge to strengthen the foundations of SE:
  produce knowledge that can be based on scientific methodology

Objectives

- Discussing the significance of experimental studies, particularly SRs and their use in supporting software processes
- Present a template designed to support systematic reviews in SE
- Introduce development of ontologies to describe knowledge regarding such experimental studies

Ontology to support systematic reviews in Software Engineering

Problem

- Major challenge to strengthen the foundations of SE:
  produce knowledge that can be based on scientific methodology

Ontology to support systematic reviews in Software Engineering

Problem

- Major challenge to strengthen the foundations of SE:
  produce knowledge that can be based on scientific methodology

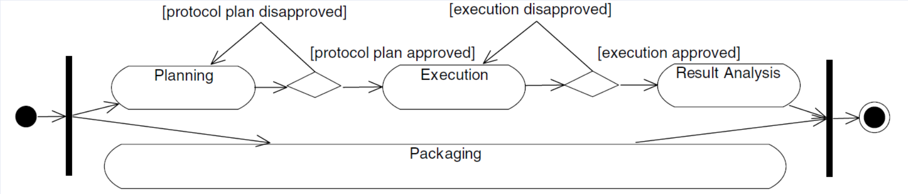
Objectives

- Discussing the significance of experimental studies, particularly SRs and their use in supporting software processes
- Present a template designed to support systematic reviews in SE
- Introduce development of ontologies to describe knowledge regarding such experimental studies

Ontology to support systematic reviews in Software Engineering

Systematic Review conduction process

1. **Planning**: research objectives and SR protocol
2. **Execution**: identify, select and evaluate primary studies
3. **Result Analysis**: extract and synthesize data from the the articles



Systematic Review protocol template

1. Question Formularization
   1.1. Question Focus
   1.2. Question Quality and Amplitude
      - Problem
      - Question.
      - Keywords and Synonyms
      - Intervention
      - Control
      - Effect
      - Outcome Measure
      - Population.
      - Application
      - Experimental Design
2. Sources Selection
   2.1. Sources Selection Criteria Definition
   2.2. Studies Languages
   2.3. Sources Identification
      - Sources Search Methods
      - Search String
      - Sources List
   2.4. Sources Selection after Evaluation
   2.5. References Checking
3. Studies Selection
   3.1. Studies Definition
      - Studies Inclusion and Exclusion Criteria Definition
      - Studies Types Definition
   3.2. Procedures for Studies Selection
   3.3. Selection Execution
      - Initial Studies Selection
      - Studies Quality Evaluation
      - Selection Review
4. Information Extraction
   4.1. Information Inclusion and Exclusion Criteria Definition
   4.2. Data Extraction Forms
   4.3. Extraction Execution
      - Objective Results Extraction
         i) Study Identification
         ii) Study Methodology
         iii) Study Results
         iv) Study Problems
      - Subjective Results Extraction
         i) Information through Authors
         ii) General Impressions and Abstractions
   4.4. Resolution of divergences among reviewers
5. Results Summarization
   5.1. Results Statistical Calculus
   5.2. Results Presentation in Tables
   5.3. Sensitivity Analysis
   5.4. Plotting
   5.5. Final Comments
      - Number of Studies
      - Search, Selection and Extraction Bias
      - Publication Bias
      - Inter-Reviewers Variation.
      - Results Application
      - Recommendations

Scientific Ontology

Suggested Ontology

Level 0: Different knowledge of domains that are involved in the conduction of SRs in SE, represented by *Experimental Method*, *Primary Research* and *Research Synthesis*

Level 1: The conceptual entity, represented by *Primary Study Element* is the highest level of hypernym of the *Primary Research* and subsumes the concepts in the lower levels of hierarchy.

Level 2: The main concepts, represented by *Structure of Study* and *Quality of Study*

Level 3: The subcategories of one of the main concept Structure of study, represented by *Problem, Hypothesis, Intervention, Control, Measurement, Outcome* and *Unit of Study*

Level 4: The entities of the subcategory *Outcome* that is demonstrating the ontological hybridism, having not only taxonomic relations, represented by *Target Outcome* and *Surrogate Outcome*, but also meronymic relations, represented by *Endpoint, Incidence, Prevalence, Effect Modification* and *Effect Modifier*

Ontology to support systematic reviews in Software Engineering

Result

- Observe: the ontology results in directly linked with Systematic review protocol template object.
- Here only the small part. The full ontology conceptualizes on all roles is SR template
- powerful, comprehensive and covers all SR needs

Ontology to support systematic reviews in Software Engineering

Result

- Observe: the ontology results in directly linked with Systematic review protocol template object.
- Here only the small part. The full ontology conceptualizes on all roles is SR template
- powerful, comprehensive and covers all SR needs

Future works

- At the moment of publication only in development
- Merging with Software Engineering Ontology and integrating into eSEE
- Towards a wider Experimental Software Engineering Ontology that will combine all received evidence-based knowledge in SE

Ontology to support empirical studies

Karlsruhe Institute of Technology

Problem

Badly designed experiment can increase the cost and risk of invalid results.

Ontology to support empirical studies

Problem

Badly designed experiment can increase the cost and risk of invalid results.

Objectives

- Present an ontology for analyzing empirical studies of SE, in particular the design of software engineering experiment
- Encapsulate the experience experts by means of an ontology for experimental designs using Protege OWL

Ontology to support empirical studies

Suggested Ontology (main concepts)

- *Treatment* Software Engineering Technique/Method/Process being studied
- *Subject* Person, Developer/Student participated on experiment
- *Object* Entity, Program/Model
- *Assignment* Relation between all of them; in an assignmentInstance: a subjectInstance is assigned to apply a treatmentInstance in an objectInstance.

Ontology to support empirical studies

Suggested Ontology (main concepts)

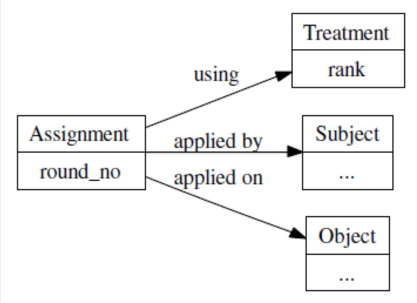


Figure: Ontology fragment depicting the concepts involved in the design of experiments (Siy and Wu 2012, p. 13)

Ontology to support empirical studies

Suggested Ontology (constraints)

1. No *Subject* can be assigned a *Treatment* that is less sophisticated than the other ones he was already assigned to. What for: subject assigned to one treatment may use the knowledge and experience gained from that treatment.
2. No *Treatment* was applied by only one *Subject*. What for: experiment subjects have varying backgrounds and abilities that implies different results. If only one subject is related, then it is not scientifically meaningful.
3. *Subject* is assigned to several *Treatments*. What for: assess the variability introduced by that subject.
4. An *Object* is treated by several *Subjects*. What for: provide a way to untangle subject performance from object complexity.

Ontology to support empirical studies

Result

- Evaluated on: experiments on software inspections (by Basili, Shull, and Lanubile 1999)
- Aligning concrete values: *Treatment*:=reading technique, *Subject*:=reviewer, *Object*:=software requirements document.
- The results are “encouraging” (Siy and Wu 2012, p. 15)
- Not found any validation errors after the applying all constraints
- If Ontology fed by invalid assignments, then observed inconsistencies as expected

Ontology to support empirical studies

Result

- Evaluated on: experiments on software inspections (by Basili, Shull, and Lanubile 1999)
- Aligning concrete values: *Treatment*:=reading technique, *Subject*:=reviewer, *Object*:=software requirements document.
- The results are “encouraging” (Siy and Wu 2012, p. 15)
- Not found any validation errors after the applying all constraints
- If Ontology fed by invalid assignments, then observed inconsistencies as expected

Future works

- Shown results: only a step towards a comprehensive ontology for other Software Engineering knowledge domains (not only the systematic review process)

Comparative Analysis

Similarities

- Adoption of ontologies: in researcher’s opninion, best for accumulate any experience or knowledge (especially, from experiments), formalize them for later representing
- Not a silver bullet: cannot depict precisely the real objects + restricted by the first-order logic.
- **But:** still enough for fulfilling a lot of objectives

Comparative Analysis

Similarities

- Adoption of ontologies: in researcher's opninion, best for accumulate any experience or knowledge (especially, from experiments), formalize them for later representing
- Not a silver bullet: cannot depict precisely the real objects + restricted by the first-order logic.
- **But:** still enough for fulfilling a lot of objectives

Differences

- Ontology for supporting systematic reviews (Almeida Biolchini et al. 2007) belongs to the area Methods
- Ontology for empirical studies (Siy and Wu 2012) belongs to as Reproducibity as Methods
- Used different ontology languages (or dialects) → barriers for applying them and making as standard

Ontology based systems in scientific search

Why?

- During the investigating of ontologies approaches: Ontologies are used in the contemporary **search engines** such as Google, Bing or Yandex.
- At the same time: **Ontology-based systems** actively used in the scientific world

Ontology based systems in scientific search

Why?

- During the investigating of ontologies approaches: Ontologies are used in the contemporary **search engines** such as Google, Bing or Yandex.
- At the same time: **Ontology-based systems** actively used in the scientific world

Question

- How are **Ontology-based systems** used in the scientific **search engines**?

Semantic Web

Problems of the current Web

- The current Web is a Web of documents, made and understandable only by humans → Semantic Web should become a Web of meaning and thereby understandable by machines
- The current Web is multi-lingual. E.g. English shopping website would use the word price to refer to an items price, Dutch - prijs, a French - prix.→ Make the different words meaning the same and thereby understandable by machines

Semantic Web

Problems of the current Web

- The current Web is a Web of documents, made and understandable only by humans → Semantic Web should become a Web of meaning and thereby understandable by machines
- The current Web is multi-lingual. E.g. English shopping website would use the word price to refer to an items price, Dutch - prijs, a French - prix.→ Make the different words meaning the same and thereby understandable by machines

W3C states: Semantic Web is about

- Common formats for interchange of data, where original Web only interchange of documents.
- Language for relations of the data to real world objects, where person/machine starting by one database, move to another, connected not only by wires but by being about the same thing,

Traditional search engine

Problems of the current Web

- Background: crawler indexes the already existing available information in WWW by storing, marking and organizing the fetched data into database, using mainly keywords
- Client gives the query
- According to keywords and many other metrics the ranked pages displayed

Traditional search engine

Problems of the current Web

- Background: crawler indexes the already existing available information in WWW by storing, marking and organizing the fetched data into database, using mainly keywords
- Client gives the query
- According to keywords and many other metrics the ranked pages displayed

Semantic search engine

- Parse the keyword and forward to the Reasoner
- Reasoner: analyze using **ontologies** and output the RDF triples
- Calculating the rank based on relations between ontologies and documents

Traditional search engine

Problems of the current Web

- Background: crawler indexes the already existing available information in WWW by storing, marking and organizing the fetched data into database, using mainly keywords
- Client gives the query
- According to keywords and many other metrics the ranked pages displayed

Traditional search engine

Problems of the current Web

- Background: crawler indexes the already existing available information in WWW by storing, marking and organizing the fetched data into database, using mainly keywords
- Client gives the query
- According to keywords and many other metrics the ranked pages displayed

Semantic search engine

- Parse the keyword and forward to the Reasoner
- Reasoner: analyze using **ontologies** and output the RDF triples
- Calculating the rank based on relations between ontologies and documents

Traditional search engine vs. Semantic one

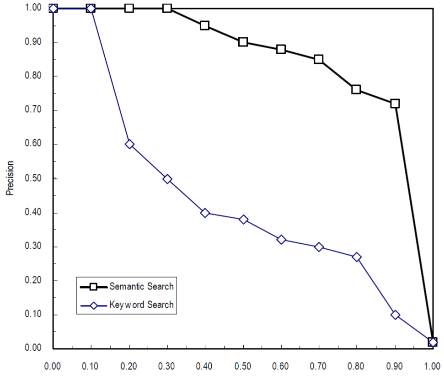


Figure: Precision-recall graph comparison of semantic and keyword searchWei-Dong Fang et al. 2005, p. 1918

Scientific ontologies

Conclusions

- Obviously: Using ontologies bring the advantages
- Problem: selecting suitable ontologies that sophisticate the initial requirements and purposes

Scientific ontologies

Conclusions

- Obviously: Using ontologies bring the advantages
- Problem: selecting suitable ontologies that sophisticate the initial requirements and purposes

Diversity of scientific ontologies

- Review by Ruiz-Iniesta and Corcho 2014.
- Classification in three main classed
- Still non-exhaustive

Scientific ontologies

Diversity of scientific ontologies

1 Describing document structure: structure of a scientific publication

- Examples of concepts: Author, Title, Volume
- Realization: simple extension for HTML-file allows web page authors to annotate their web documents with machine-readable knowledge
- Developed in pre-semantic era
- Example: Document Ontology

Scientific ontologies

Diversity of scientific ontologies

1. Describing document structure: structure of a scientific publication
   - Examples of concepts: Author, Title, Volume
   - Realization: simple extension for HTML-file allows web page authors to annotate their web documents with machine-readable knowledge
   - Developed in pre-semantic era
   - Example: Document Ontology

Example for Journal Future-Generation-Computer-Systems

author(Paper1-FGCS, Daniel-Garijo)
title(Paper1-FGCS, “Motifs in Scientific Work Empirical Analysis”)
volume(Future-Generation-Computer-Systems, 123)

Scientific ontologies

Diversity of scientific ontologies

2. Describing the rhetorical elements: rhetorical structure of a publication independent on any research field
    - Consist of 3 parts: **Header, Body, Tail**
    - **Header** concepts: Creator, Title etc., **Body**: Introduction, Method etc.; **Tail**: Acknowledgment, Reference etc.
    - Example: Ontology of Rhetorical Box (uses RDF)

Scientific ontologies

Example for Journal Future-Generation-Computer-Systems

@prefixdc :< http : //purl.org/dc/elements/1.1/> .

@prefixdcterms :< http : //purl.org/dc/terms/> .

@prefixorb :< http : //purl.org/orb/>

[dc : creator" Garijo, Daniel" ;

dc : title" Commonmotifsinscientificworkflows..." ;

dcterms : abstract" Workflowtechnologycontinuestoplay..."

dcterms : hasPart

[aorb : Header; dc : description" Headercontent"],

[aorb : Introduction; dc : description" Introductioncontent"]

...

[aorb : References; dc : description" Listofreferences"]

Scientific ontologies

Diversity of scientific ontologies

3 Describing bibliographies and cites structure: citating and referencing of a scientific publication

- Examples of concepts: Author, Citation, Institution
- Realization: OWL DL
- Example: FaBiO (publishing on the Semantic Web bibliographic records of scholarly endeavours.) and CiTo (characterization of citations, both factually and rhetorically)

Scientific ontologies

Diversity of scientific ontologies

3 Describing bibliographies and cites structure: citating and referencing of a scientific publication

- Examples of concepts: Author, Citation, Institution
- Realization: OWL DL
- Example: FaBiO (publishing on the Semantic Web bibliographic records of scholarly endeavours.) and CiTo (characterization of citations, both factually and rhetorically)

Example

$@prefix :< http : //www.sparontologies.net/example/> .@prefixcito :< http : //purl.org/spar/cito >$
$: citationacito : Citation;$
$cito : hasCitingEntity : paper - a;$
$cito : hasCitationCharacterizationcito : extends;$
$cito : hasCitedEntity : paper - b.$

Scientific ontologies

Other ontologies

- Argumentative ontologies with concepts like claims, constructive and comparative statements about other works, etc.
- Domain-specific ontologies e.g. Mathematics ontology with concepts Stetement, Proof, etc. or Computer Science Ontology

Scientific ontologies

Other ontologies

- Argumentative ontologies with concepts like claims, constructive and comparative statements about other works, etc.
- Domain-specific ontologies e.g. Mathematics ontology with concepts Stetement, Proof, etc. or Computer Science Ontology

Computer Science Ontology

- “Automatically generated ontology of research areas, which includes about 14K topics and 162K semantic relationships” (Salatino et al. 2019, p.1),
- Main purpose: taxonomy for the Computer Science e.g. Deep Lerning is a part of Machine Learning

Requirements for scientific ontologies

Knowledge Graph

- The main skeleton of scientific search engines: **Knowledge Graphs** (KGs)
- “(1) an ontology describing a conceptual model, and (2) the corresponding instance data following the constraints posed by the ontology” (Brack et al. 2020, p. 1)
- Created by different approaches
  - Automatic construction from text
  - Information extraction from scientific text (NLP)
  - Manual curation

Requirements for scientific ontologies

Knowledge Graph

- The main skeleton of scientific search engines: **Knowledge Graphs** (KGs)
- “(1) an ontology describing a conceptual model, and (2) the corresponding instance data following the constraints posed by the ontology” (Brack et al. 2020, p. 1)
- Created by different approaches
  - Automatic construction from text
  - Information extraction from scientific text (NLP)
  - Manual curation

Examples of KGs

- Wikipedia, Google Scholar and Microsoft Academic with corresponding and WikiData KG, Microsoft Academic Knowledge Graph and SciGraph

Requirements for scientific ontologies

Use cases for KGs

- getting research field overview
- finding related work
- assessing relevance
- extracting relevant information
- getting recommended articles
- obtaining deep understanding
- reproducing results

Requirements for scientific ontologies

Karlsruhe Institute of Technology

Non-functional requirements for KGs

- domain-specification (e.g. high in CSO)
- granularity (e.g. how circumstantial)
- coverage (e.g. number of scientific fields)
- quality (e.g. reliability of the instance data)

Requirements for scientific ontologies

Non-functional requirements for KGs

- domain-specification (e.g. high in CSO)
- granularity (e.g. how circumstantial)
- coverage (e.g. number of scientific fields)
- quality (e.g. reliability of the instance data)

Conclusion

- Combine use cases and non-functional requirements in 4 main groups

Requirements for scientific ontologies

Main classes of ontologies

1. High domain-specification, granularity, quality and low coverage (e.g. maintaining of the structure, evolution and relevance of contained information)
2. High coverage, low domain-specification, granularity, quality (e,g, searching for related work)
3. High quality, medium domain-specification and granularity, low coverage (e.g. linking or representing of semantic description such as guidelines or standards)
4. Medium domain-specification, coverage, quality and low granularity (e.g. highlighting zones))

Requirements for scientific ontologies

Main classes of ontologies

1. High domain-specification, granularity, quality and low coverage (e.g. maintaining of the structure, evolution and relevance of contained information)
2. High coverage, low domain-specification, granularity, quality (e,g, searching for related work)
3. High quality, medium domain-specification and granularity, low coverage (e.g. linking or representing of semantic description such as guidelines or standards)
4. Medium domain-specification, coverage, quality and low granularity (e.g. highlighting zones))

Recommendations

- Manual: 1. class, Automatic: 2. class
- Hybrid (automatic, then manual): 3. and 4. classes

Scientific ontologies without KG integrations

Important

- Not only approaches with KGs
- Also independent solutions

Scientific ontologies without KG integrations

Karlsruhe Institute of Technology

Important

- Not only approaches with KGs
- Also independent solutions

Independent approaches

- Ontology based Semantic Search Engine for Cancer: analysis of cancer, its categories, types, causes, symptoms, etc. (Raj and Sarumathi 2014)
- Retrieving information from scientific abstracts (Milward et al. 2005)
- Ontology Web Search Engine: automatic search and merging of existing ontologies → terminal purpose: Semantic Web Expert System (SWES) - able to give answers like an expert for different types of queries, in particular scientific ones (Verhodubs 2015)

Conclusions

Ontologies

- The best tool for interchanging of pure information independent on languages, definitions and other syntactic barriers
- Effectively reuse and standardize of the obtained knowledge
- Contemporary ontologies based on strictly defined in mathematical logic ontology languages

Conclusions

Ontologies

- The best tool for interchanging of pure information independent on languages, definitions and other syntactic barriers
- Effectively reuse and standardize of the obtained knowledge
- Contemporary ontologies based on strictly defined in mathematical logic ontology languages

Meta-research

- Research on research
- how researches should be conducted, what practices effective and in what fields
- Diversity of meta-research research: Methods, Reporting, Reproducibility, Evaluation, Incentives.

Conclusions

Karlsruhe Institute of Technology

Ontologies in Software-Engineering-Meta-Research

- Huge contribution to Evidence-based Software Engineering
- Support determining the best SE practices using systematic reviews in secondary studies
- Useful for packaging of controlled experiments in primary studies

Conclusions

Karlsruhe Institute of Technology

Ontologies in Software-Engineering-Meta-Research

- Huge contribution to Evidence-based Software Engineering
- Support determining the best SE practices using systematic reviews in secondary studies
- Useful for packaging of controlled experiments in primary studies

Ontologies in scientific semantic search engines

- The skeleton of any semantic search engine
- Transition from the Web of documents to web of meaning
- Accelerates and improves the search process for scientists

References I

- Jorge Calmon de Almeida Biolchini et al. "Scientific research ontology to support systematic review in software engineering". In: *Adv. Eng. Informatics* 21.2 (2007), pp. 133–151.
- V. R. Basili, F. Shull, and F. Lanubile. "Building knowledge through families of experiments". In: 25 (1999), pp. 456–473. ISSN: 0098-5589. DOI: 10.1109/32.799939.
- Arthur Brack et al. "Requirements Analysis for an Open Research Knowledge Graph". In: (May 20, 2020). arXiv: 2005.10334v1 [cs.DL].
- John P. A. Ioannidis et al. "Meta-research: Evaluation and Improvement of Research Methods and Practices". In: 13 (), e1002264. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1002264.
- Boris Konev. *Lecture notes of course Ontology and knowledge representation*. 2010. URL: https://www.lektorium.tv/speaker/2680.
- David Milward et al. "Ontology-Based Interactive Information Extraction From Scientific Abstracts". In: 6 (2005), pp. 67–71. ISSN: 1531-6912. DOI: 10.1002/cfg.456.
- Syam Raj and S Sarumathi. "Ontology based Semantic Search Engine for Cancer". In: *International Journal of Computer Applications* 95.5 (2014), pp. 39–43. ISSN: 0975-8887. DOI: 10.5120/16594-6308.

References II

- Almudena Ruiz-Iniesta and Oscar Corcho. “A review of ontologies for describing scholarly and scientific documents”. In: *Proceedings of the 4th Workshop on Semantic Publishing co-located with the 11th Extended Semantic Web Conference (ESWC 2014), Anissaras, Greece, May 25th, 2014*. Ed. by Alexander Garcia Castro et al. Vol. 1155. CEUR Workshop Proceedings. CEUR-WS.org, 2014. URL: `http://ceur-ws.org/Vol-1155/paper-07.pdf`.
- Angelo Salatino et al. “The Computer Science Ontology: A Comprehensive Automatically-Generated Taxonomy of Research Areas”. In: (2019). URL: `http://oro.open.ac.uk/66268/1/DataIntel-CSO.camera.ready.ORO.pdf`.
- Harvey Siy and Yan Wu. *An Ontology to Support Empirical Studies in Software Engineering*. en. 2012. URL: `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.216.8899;%20http://csalpha.ist.unomaha.edu/~hsiy/research/icc09.pdf`.
- Olegs Verhodubs. “Towards the Ontology Web Search Engine”. In: (May 4, 2015). arXiv: `1505.00755v1 [cs.IR]`.
- Wei-Dong Fang et al. “Toward a semantic search engine based on ontologies”. In: *2005 International Conference on Machine Learning and Cybernetics*. Vol. 3. 2005, 1913–1918 Vol. 3.