Innovation and Research, discussion on literature management

Alexandre Benoit

Université Savoie Mont-Blanc, LISTIC, Annecy, France, 2023

Course outline

Course Objectives and Organization

Context

The Bibliographic Sources and Information Validation, an Introduction

Course Objectives and Organization

Objectives

The objective of this module is to give the methodological bases on:

- Literature search.
- Bibliographic synthesis, i.e. the report and critical analysis of a set of documents on the same theme, based on explicit criteria.
- Paper reviews, i.e. how to estimate paper quality and the proposed contributions.
- Improve research visibility and encourage open science.

Course schedule and organization

8h lesson with some practice:

Lesson	Volume	Objectives
CM 1	2h	General presentation, context, goals and means.
CM 2	2h	bibliographic references management : collect, organize, integrate.
CM 3	2h	Conducting literature searches: resources, strategies, plagiarism.
CM 4	2h	Promoting your research: digital identity, networking, bibliometrics.

Teachers (contact : firstname.lastname@univ-smb.fr)

- Alexandre Benoit, Professor at LISTIC.
- Gaëlle Charra& Christelle Serra, University library manager at Bourget du lac.
- Pauline Simon&Audrey Stefani, University library manager at Annecy, instructor.
- Julie Alibert, University Open science advisor.

Context

Context

Why and how?

Why is literature survey mandatory ?

- Literature survey and reporting is a fundamental activity of scientific work.
- Searching for information, making a structured synthesis and critic provides an overview of state of the art in a given domain.
- **This is a mandatory step to position a contribution and a novel approach** with respect to state of the art.

Something to know: the scientific knowledge dissemination process

Many workflows exist, here are some typical examples:

- A laboratory/a research team proposes a contribution → Scientific publications are published → Knowledge spreads → Popular media propose simplified summaries.
- A company proposes a contribution → A patent is published → contribution spreads → Popular media propose simplified summaries.
- ...

Knowledge confidence level across the workflows and associated pricing (approx.):

Expert level, $\sim 1000\$/day$

↓

Scientific publications, $\sim 100\$$**/paper**

↓

Popular media, low precision-level description, poor validation processes, cheap/free

Context

Scientific documentation document types

Document types overview vs confidence levels

A general picture:

Type	Provided details	Confidence level	Topics area
PhD Thesis	++	++	"very" wide
Scientific Journal article	++	++	wide
Scientific Conference Article	+	+	narrow
Reports (master, technical)	+	+	narrow
Patent	+	+	narrow
Periodicals (specialized)	$\sim$	$\sim$	maybe wide
Web article	$\sim$	$\sim$	maybe wide
Social network article	$\sim -$	$\sim -$	maybe wide
(Specialized) blog article	$\sim -$	$\sim -$	maybe wide
Official documents	++	+++	maybe wide

WARNING : this is only a general picture. Actually, for each category, a strong variability in the quality and confidence level occurs and also depends on the research domain !

Special note on Patents

A variety of spatial limitations (national, European, etc.) with temporal boundaries.

The aim is to protect intellectual property.

→ often **few details are provided**, more general information is provided in order to extend the patent boundaries.

→ always try to find related scientific publications associated with the patent that may provide more details. Find papers from the associated research teams or challenger papers.

Where to find and explore patents:

- National scale: National Institute of Industrial Property (INPI).
- Worldwide scale : World Intellectual Property Organization (WIPO)

Special note on academic publications

For either journal articles and conference papers, here is a typical structure:

- Abstract (main ideas and contributions in a few words).
  → **will facilitate paper selection for your literature surveys** !
- Introduction (context, global positioning).
- State of the art / literature survey: pros and cons of existing solutions that help defend the proposed contribution.
- Method presentation.
- Evaluation.
- Conclusions.
- References.

Paper quality depends on: The journal and conference that publishes the paper. There quality is constantly evaluated, check out tools such as CORE Rankings portal and SJR (next lessons will provide more details).

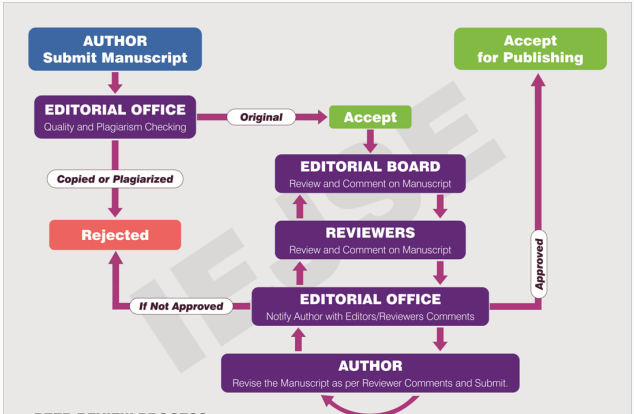
Use arXiv papers with care ! Always check if a paper of interest has finally be published in a conference/journal of enough quality.

The Bibliographic Sources and Information Validation, an Introduction

Confident sources, papers with a review process

Always look for papers with a review process ! Editors like IEEE, Elsevier, ACM and so on imposing several constraints that favour paper quality... thus making them confident sources.

Here is a typical validation process (from `http://iejse.com/peer-review-process.html`):



Confident sources

For scientific journal and conference papers, the review process quality impacts on the ranking reported on CORE, **SJR**... Have a look at it !

Scientific paper portal for French researchers is HAL. This can be considered a confident source. Researchers are asked to classify their published and unpublished papers and to upload them on this platform that then feeds arXiv automatically.

Once again, many arXiv papers are not reviewed... take the time to check their publication acceptance at journals and conferences.

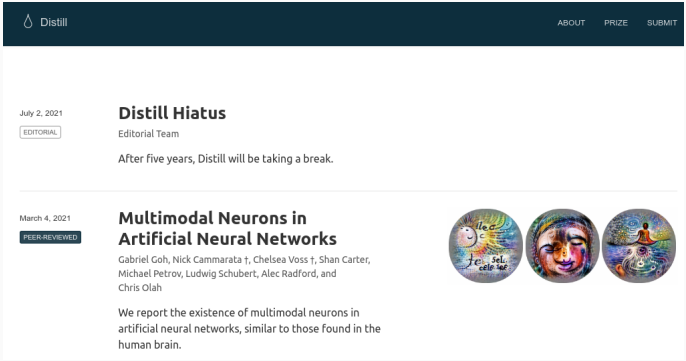
Additional general rules for confidence evaluation:

- Have a look at the authors, editors.
- Check if the publication date complies with your studies.
- Check source domain (URL and sources aim).
- Source popularity, ranking.
- Check content quality (structure, presentation and arguments quality).

Online ressources

Regarding online resources (blogs, articles, etc.?), few of them are reviewed. For instance, Towardsdatascience and Medium are well known but the related review processes if any is not as strict as for published scientific papers.
→ **But** those sources can be an entry point to look for reviewed sources !

Also, some specific web platforms with review processes exist such as distill.pub... but their economic model is sensitive.



Assignement

To be used in the next sessions and to encourage communication with your PhD advisor(s) on the publication topic:

Identify with your advisor(s):

- 5 journals,
- 5 conferences,

that would be appropriate for state of the art monitoring and publication of your own work.

You should also discuss about their ranking. Note that some conferences can be higher ranked than others but some are more interesting than others regarding collaboration opportunities to build up projects and so on.

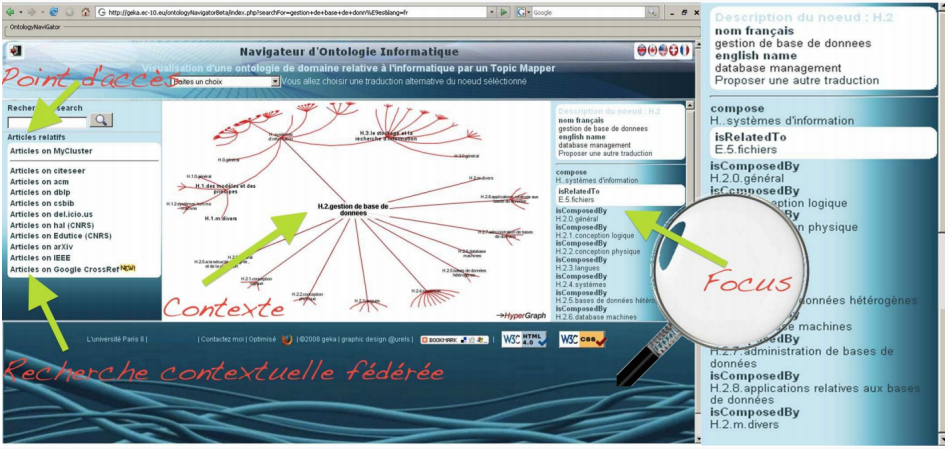
This assignment will serve as a basis for the next lessons 3 (literature survey) and 4 (communication).

Tools

One needs help !

Literature is huge, manually looking for information is difficult. Exploration tools design and evaluation is a research topic that embeds neuroscience, data visualization, computer science and so on !

An example, Kembellec phD (2012) [Kem12] presents OntologyNavigator



Online literature search tools

Meta Aggregators

- Google Scholar. A now classical tool that provides search tools, metrics, personal libraries, and so on.
- Base
- Scientific paper portal for French researchers is HAL with search tools.
- ...

Systems with personal accounts and recommendation tools can be helpful... (Google Scholar, and others).

Desktop (connected) tools

One needs to store collections of references. To grab them, web browser extensions are available and helpful !

Some classical extensions associated with specific local/online applications:

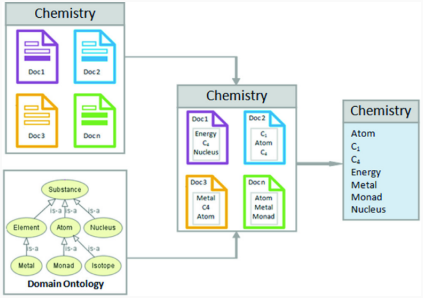
- Mendeley from Elsevier.
- **Zotero**, **open-source !**

Desktop solutions:

- Docear, open source, complete suite with a recommendation system, **MindMapping representation**... really of interest but not really supported now.
- JabRef, low-level bibtex reference manager.
- Mendeley and Zotero desktop applications.

How to organize articles databases ?

A good practice is to follow an ontology guided approach as described in [IK16; Kem12]:



from [IK16]

However domain ontologies are not always available and are subject to change. Popular systems such as Mendeley and Zotero allow articles to be associated with (personalized) keywords. At least use this to construct a literature tree or graph. If available, rely on domain ontologies to facilitate database sharing !

Domain ontology

From [IK16]: It represents concepts for describing a domain and interpreting a description of a problem in that domain. A 5-tuple based structure can be considered to describe the concepts and their relationships of a particular domain.

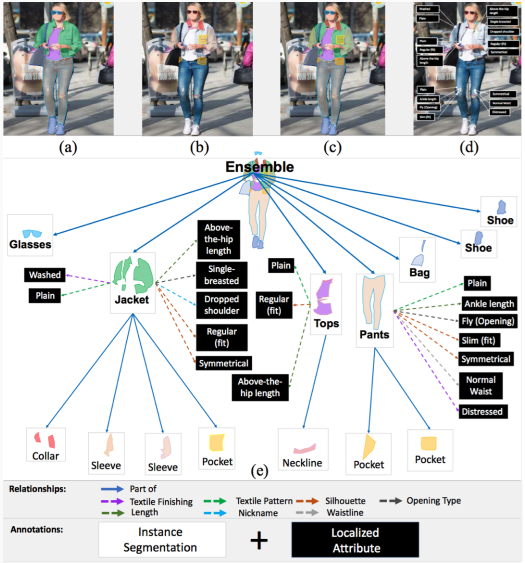
$$D = (C, I, H, type, rel) \tag{1}$$

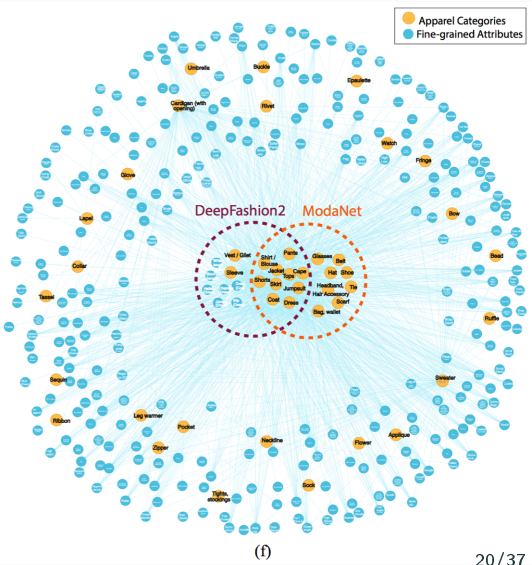
where:

- C is a finite set of concepts.
- I is a finite set of lexical entries (Instances).
- H is a finite set of concept to concept relationships.
- $type$ is a finite set of instances to concept relationships.
- rel is the finite set of instance to instance relationships.

Domain ontology visual illustration : FashionPedia

Alexandre Benoit





Computer Science Domain Ontology

A first example From [LDY11] (2011)

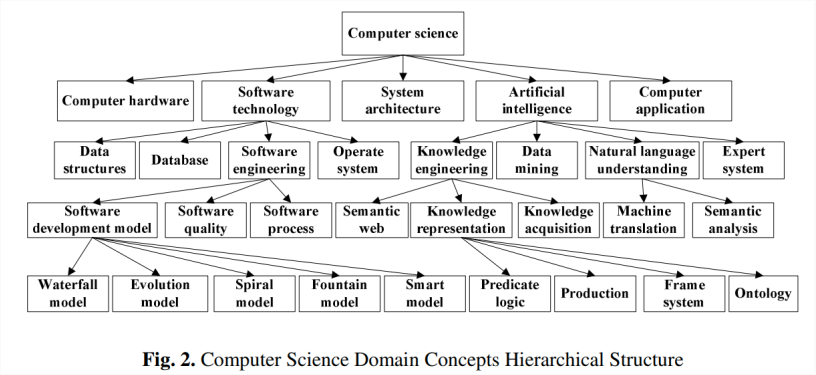
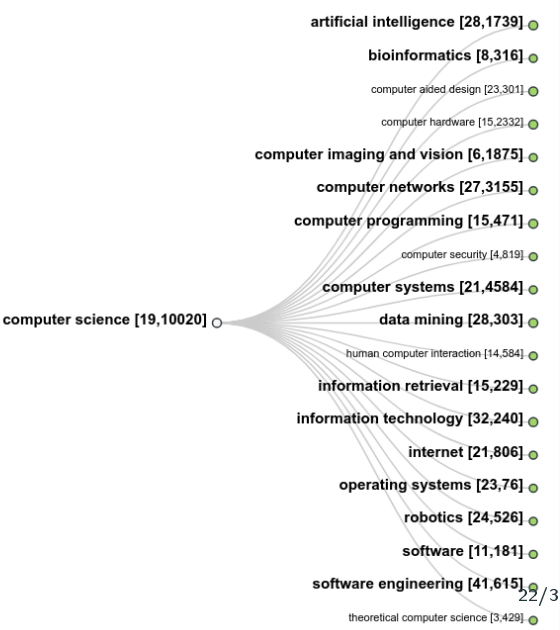


Fig. 2. Computer Science Domain Concepts Hierarchical Structure

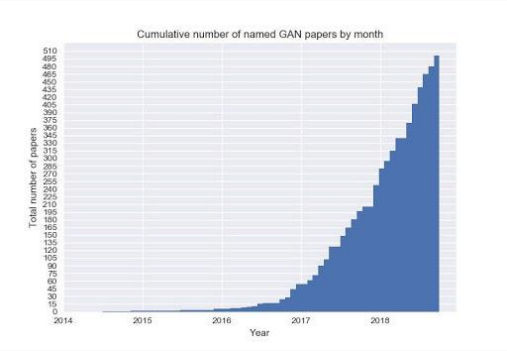
Computer Science Domain Ontology

A second example from [Sal+18] (2018).
Web interactive application :
`http://cso.kmi.open.ac.uk/topics/computer_science`
→ *different ontologies of the same domain can then be defined and can evolve in time !*



Literature monitoring

Stay up to date with recent literature... but don't get lost and don't panic !



Online tools come to the help.

→ Scientific editors and vulgarization platforms help.

→ recommender systems provide more focused alerts. Often included in tools such as Zotero, Mendeley, GoogleScholar, and so on.

→ *ChatGPT like tools... to be used with care and ethics... may generate false references.*

Literature survey sources and tools, some conclusions

- Literature even on a specific domain is rich.
- Focus on scientific papers.
- Rely on scientific paper search system and keep an eye on science vulgarization platforms to redirect to scientific papers of interest.
- Rely on specific scientific literature database systems to store your publications of interest, notes and so on.
- Organize your literature following along structured representations (ontologies when applicable).

Writing a literature survey

Conducting your own survey (at least your first year task)

Along the first year of your PhD, you are expected to:

- Explore and understand the research on your topic.
- Refine scientific questions on your topic and adjust your PhD directions.
- ...

Building a literature survey thus makes sense !

Literature survey, discourse orientation

Where and how to start ?

→ Identify a few central questions that cover the scope of your studies. Refine them with secondary (personal/application guided) questions that will help refine the investigations.
→ identify the application domain(s) and the related scientific domain(s).
→ Find out "*survey* papers", a specific category of journal papers that provide a wide overview on a topic with a very rich literature survey. This will help refine your questions and positioning.

Literature survey or review ?

Some terminology... you may distinguish:

- **Surveys**: provide an overview on a scientific question. Collects and describes many references but few opinions, refined analysis and search directions are reported...
  your first month of work for the PhD?
- **Reviews/narrative review**: more extensive collection of references, more analysis, opinions provided... In depth comparison of the different studies, with conclusions on strength and weakness. It explicitely presents research directions. Longer and more qualitative paper...
  your first phD year result ?
- **Systematic review**: extended review with a (transparent) systematic strategy. Includes unpublished(gray) papers. Makes use of robust data extraction strategies. Provides refined analysis and statistics on the research directions and contributions...
  generally >1 year work.

Literature survey organisation

Many approaches are possible but think about the reader and your own representation of the problem !
→ A coarse to fine approach can be adopted to gradually focus from the general context to the specific topic of interest.

→ Chronological presentation is an approach but not always the most appropriate to present the whole topic. It is, however, often useful to detail the evolution of a specific segment/axis of the presentation.

→ Hihllight relations between the different concepts (ontologies can help).

References reporting

Once text with citations is written, how to list the considered references in the bibliography section ? A variety of systems have been proposed, bibliography style generally chooses among:

- Alphabetical order (APA norms).
- Importance with primary sources shown first.
- Chronological order (Harvard system)
- Appearance ordering (Vancouver system): same order as it appears within the discourse.
- Thematic organization, more structured.

→ Will be experimented along lessons 2 and 3.

Checklist for literature reporting

A few recommendations from Scribbr.fr with some fine-tuning:

- Report literature in a logical manner.
- **each reported reference main idea MUST be provided/explained !** Then,
  - Limit to reliable information.
  - Avoid massive reference lists. The reported references are all related to the subject of the article.
- Style must be homogeneous across references, templates exist ! example, IEEE.
- Online references (websites, online articles) must be available.
- Avoid unexplained abbreviations.

A note on plagiarism

Plagiarism is a fault, pay attention...
Consequences can be... important for you and all the co-authors, labs, universities and industrial partners !
Some details here.

Journals may apply plagiarism detection tools before any reviewing... to limit problems at an early stage. But papers with plagiarism may be published and retroactive sanctions can be applied !

Never forget to cite papers that helped your understanding of the topic.

Connexions with the ethics module lessons !

A note on (auto)plagiarism

Some typical situations to pay attention to:

- classical plagiarism issue :
  - authors forget to cite work of other authors.
  - authors cite other work but do not report appropriately (simple copy/past of sentences without rephrases, nor *"quotes"*)
- an author describes the same contribution on 2 different publications (auto-plagiarism).
  - → general rule : a new paper must introduce new contributions but can build on a previous one. New contribution must be highlighted and previous paper must be cited !
  - → border line situation : a paper overlaps with the PhD manuscript content.
- Making use of generative models (ChatGPT and others): original texts without references may be proposed... **CHECK SOURCES !**

Connexions with the ethics module lessons !

Reviewing Articles

Is a paper relevant for reviewing ?

"I consider four factors: whether I'm sufficiently knowledgeable about the topic to offer an intelligent assessment, how interesting I find the research topic, whether I'm free of any conflict of interest, and whether I have the time. If the answer to all four questions is yes, then I'll usually agree to review."
- Chris Chambers, professor of cognitive neuroscience at Cardiff University in the United Kingdom.
from `https://www.science.org/content/article/how-review-paper`

Paper analysis

- Rely on grid analysis
  - "Standard grids" as proposed at the University Library.
  - Following some specific guidelines (conference/journal/doctoral school and so on).
- Carefully read the target paper and take notes.
- Identify the main contribution and how it is defended.
- identify the potential leaks and **try to propose improvements**:
  - From paper quality, writing quality up to structure enhancements.
  - On the contribution positioning with respect to state of the art. Check if important references are missing and could mitigate the novelty presented in the paper.... a quick Google Scolar search often helps !
  - On the contribution detailed presentation quality/accessibility.
  - On the experimental validation and proposed conclusions.

Some timing ideas: a review takes from 1 hour to a day for good papers ! 4 pages/hour is not uncommon. Then take your time for a good review !

Typical review form

From `http://alexpetrov.com/memes/sci/review_form.html`

```
                REVIEW FORM
               ------------

Author:           . . . . . . . . . .
Title of Paper:   . . . . . . . . . .

Decision  (please return within 3 weeks)
 [ ] Accept as is
 [ ] Accept with minor revisions
 [ ] Accept with major revisions (re-review)
 [ ] Reject but encourage re-submission after the work is more developed
 [ ] Reject


Rate the following items (on a scale of 1 to 10, with 10 being the best)

Significance:
 - How important is the work reported?  Does it attack an
     important/difficult problem (as opposed to a peripheral/simple one)?
 - Does the approach offered advance the state of the art?
 - Does it involve or synthesize ideas, methods, approaches from multiple
     disciplines?
 - Does it have interesting implications for multiple disciplines?

Originality:
 - Is this a new issue?  Is this a novel approach to an issue?
 - Is this a novel combination of familiar ideas/techniques/methods/approaches?
 - Does the paper point out differences from related research?
 - Does the paper properly situate itself with respect to previous work?

Quality:
 - Is the paper technically sound?  How are its claims backed up?
 - Does it carefully evaluate the strengths and limitations of its contribution?

Clarity:
 - Is the paper clearly written?  Does it motivate the research?  Does it
     describe clearly the methods employed (e.g., experimental procedures,
     algorithms, analytical tools), if any?
 - Are the results, if any, described and evaluated thoroughly?
 - Is the paper organized in a sensible and logical fashion?

Relevance:
 - Is the paper closely related to the theme of the journal (broadly conceived)?
 - Is the content interesting enough to a broad audience?
 - Is the paper readable in a multi-disciplinary context?


Explanations for the above ratings and other general comments on major
issues (on a separate sheet)


Comments on the minor details of the article (optional; on a separate sheet)
```

Writing the review

Maybe follow a standardized form. Some classical items:

- Always briefly present the paper : domain, proposed contributions, applications/experiments.
- Provide a brief summary of the proposed contribution to show what you as a reader understood.
- Then, detail your review.
- Evaluate review items to get a coarse to fine report on the paper quality.

Reviews and your PhD...

Note that similarly to the conference, journal and maybe book chapters you will write along your PhD, your final manuscript will be reviewed. Reviewers will assess your final manuscript quality and relevance and will allow (or not) for the defence to be conducted. It will also have an impact on your academic career.

$\rightarrow$ your PhD advisors will provide you some help on your final manuscript redaction but you are the author and they do not provide the final assessment. They will guide you along the PhD, tell you when the defence is possible and finally will show you directions to help enhance the contributions in the report and final presentation... hard work...

As a final lecture, an interesting article [GSL14]: "What examiners do..."

References

[Sal+18] Angelo A. Salatino, Thiviyan Thanapalasingam, Andrea Mannocci, Francesco Osborne, and Enrico Motta. "The Computer Science Ontology: A Large-Scale Taxonomy of Research Areas". In: *The Semantic Web – ISWC 2018*. Ed. by Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Aimée Kaffee, and Elena Simperl. Cham: Springer International Publishing, 2018, pp. 187–205. ISBN: 978-3-030-00668-6.

[IK16] Ali Shariq Imran and Zenun Kastrati. "Pedagogical Document Classification and Organization Using Domain Ontology". In: *Learning and Collaboration Technologies*. Ed. by Panayiotis Zaphiris and Andri Ioannou. Cham: Springer International Publishing, 2016, pp. 499–509. ISBN: 978-3-319-39483-1.

[GSL14] Clinton Golding, Sharon Sharmini, and Ayelet Lazarovitch. "What examiners do: What thesis students should know". In: *Assessment & Evaluation in Higher Education* 39.5 (2014), pp. 563–576.

[Kem12] Gérald Kembellec. “Bibliographies scientifiques : de la recherche d’informations à la production de documents normés”. Theses. Université Paris VIII Vincennes-Saint Denis, Dec. 2012. URL: https://tel.archives-ouvertes.fr/tel-00771553.

[LDY11] Dandan Li, Jianwei Du, and Shuzhen Yao. “Research on Computer Science Domain Ontology Construction and Information Retrieval”. In: *Knowledge Engineering and Management*. Ed. by Yinglin Wang and Tianrui Li. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 603–608. ISBN: 978-3-642-25661-5.