How to work with scientific literature?

How to prepare literature reviews?

How to learn from literature sources?

Ziemowit Dworakowski

AGH University of Krakow

Ranking of knowledge sources

Scientific sources

Science popularization sources
- professional webpages (e.g. NASA)
- University webpages
- Science-popularization blogosphere

Specialized internet fora

General internet services

General fora, comments, facebook, conspiration theories, etc.

- Well-prepared, highly-cited articles from high-ranked journals (**JCR list, high IF**) written by authors with high **h-index**

- Highly cited books prepared by recognized authors

- Other articles from high-ranked journals

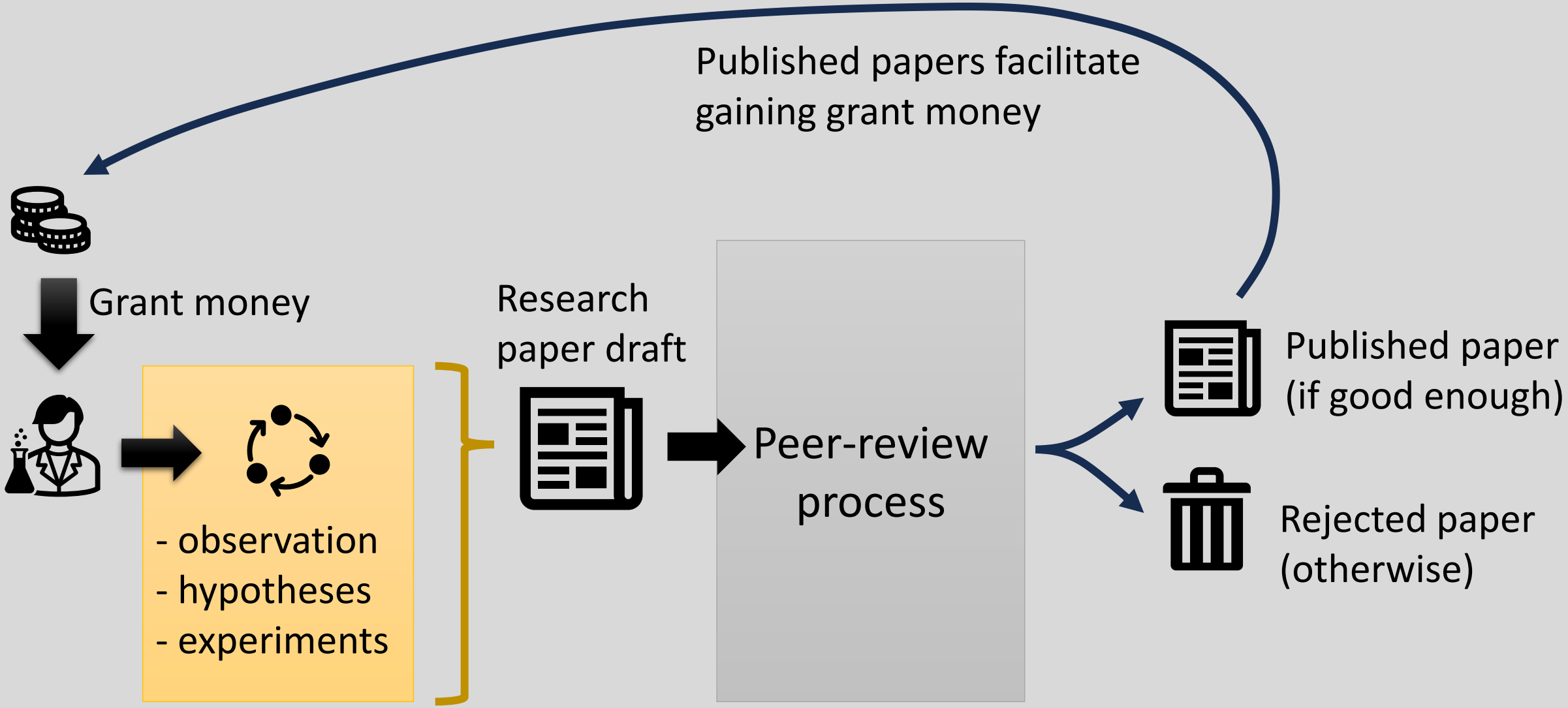
- Articles in „normal” journals

- Conference proceedings

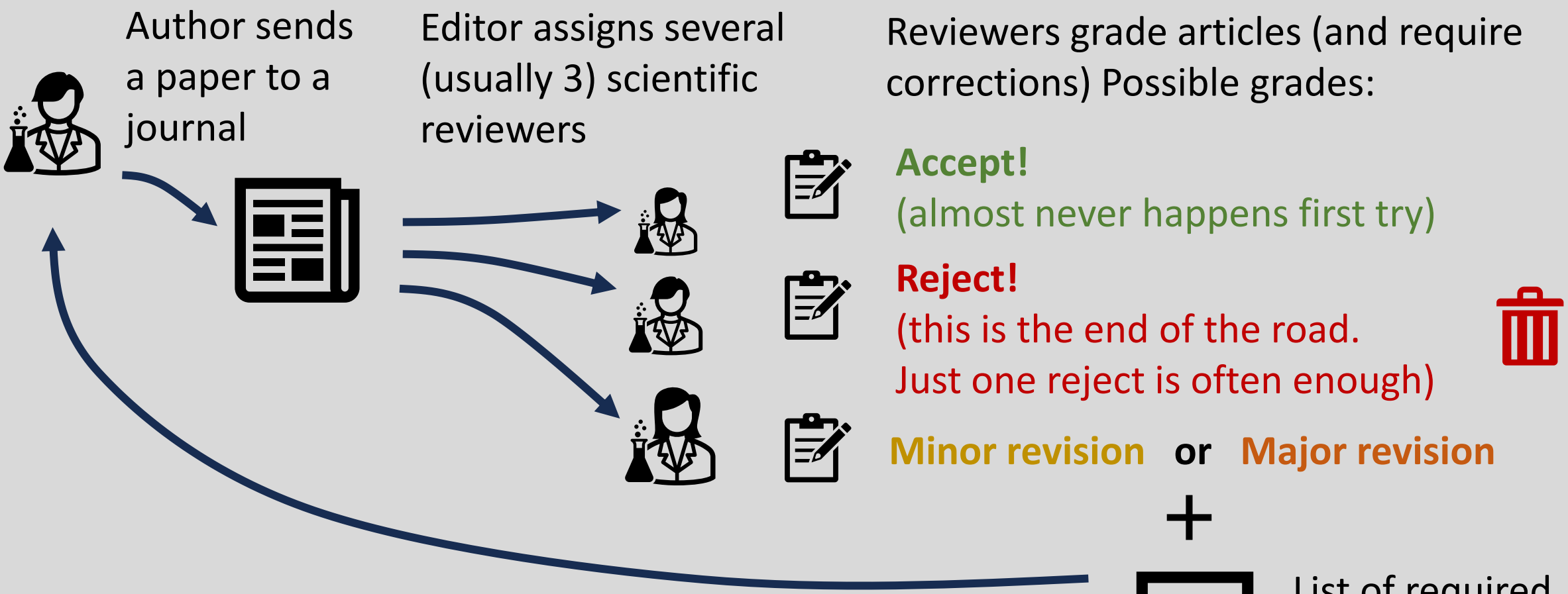
- Typical „scientific books” and doctoral theses

- Master theses and technical reports afiliated on universities (the so called: „white papers”)

How does science work?



What is inside the peer-review process?



Author uses reviews to improve the article and submit it again.
This is iterative process, Usually 2, sometimes more cycles.
Article is published **if all reviewers say „Accept!"**

What kinds of papers are there?

Research paper

- A „typical” paper
- Contains „novelty” (new finding, new method, new hypothesis, new solution)
- Usually pursues one goal (shows only such results as to prove the point)

Case study

- Usually longer than typical research paper
- Contains in-depth analysis of a particular application case
- Usually shows constraints, thought process and results of sub-decisions

What kinds of papers are there?

Conference paper

- Short paper that accompanies a conference presentation
- Usually contains a proof-of-concept of a method (example of working method with no details)
- Rarely subjected to a strict peer-review
- Good as example of application, poor choice for method understanding

Review (or survey)

- Meta-analysis of many papers, often contains hundreds of references
- Reports state-of-the-art in a predefined area
- Does not contain „new stuff” but rather combines findings of many articles
- Often the best place to start learning about a new area

Other sources of knowledge

Didactic books

- Contain a well-organized structure and complete knowledge on topic

- Prepared to facilitate learning in a particular area (often for students or young researchers)

- Often written by scientists with an established position in the field

„Review" books

- Contain large numbers of examples and references

- A different version of a review paper

- Often written by scientists for the purpose of gaining academic degrees

White papers

- Project or company reports

- Sometimes follow a „tutorial approach" and include basic explanations

- Sometimes assembled by competent authors and can be useful

Often not strictly reviewed. Can be a good source of knowledge, but usually require either your own expertise or external recommendation

Are all the sources similar?

Each journal aims to publish the best-quality research.

But each journal needs to publish particular number of papers each month.

If a journal do not have enough good papers, it will publish just decent ones

Each scientist aims to publish papers in the best possible journals

But each scientist needs to publish at least some papers each year

If finished research is not good enough, it will just be send to a „lower quality” journal

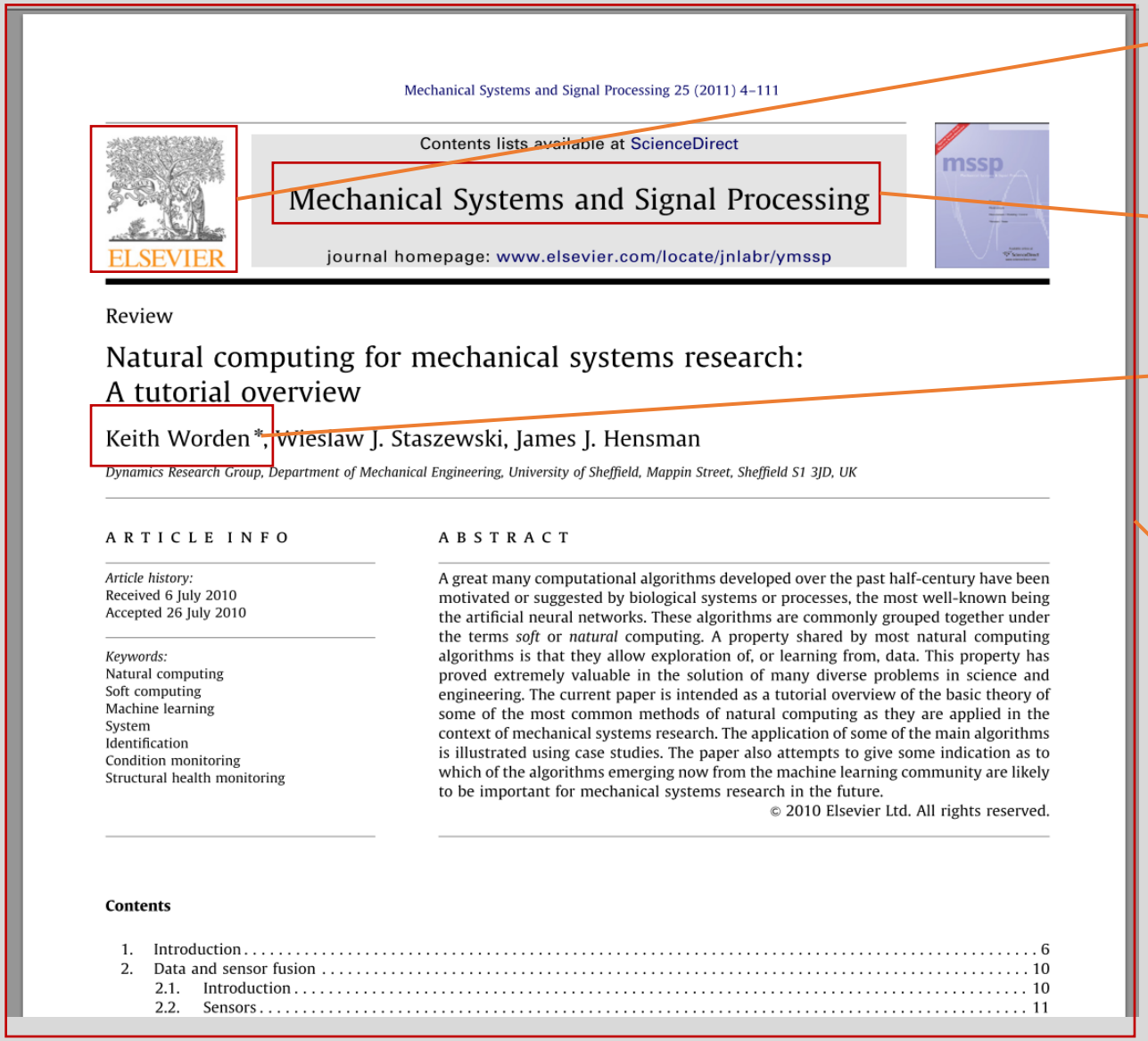
Quality metrics

Number of citations → Measures **paper quality**. Tells us how many sources refer to this paper (note: this parameter increases over time!)

IF - Impact Factor → Measures **journal quality**. Tells us what is the number of citations on average per paper within a 5-year time window

h-index (Hirsh index) → Measures **author quality**. If autor's h-index = ***n***, then at least ***n*** papers of this author were at lest ***n*** times cited. (note: this parameter increases over time! H-index should be roughly equal to research experience).

How to recognize a good paper?



Mechanical Systems and Signal Processing 25 (2011) 4–111

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/jnlabr/ymssp

Review

Natural computing for mechanical systems research: A tutorial overview

Keith Worden *, Wiesław J. Staszewski, James J. Hensman

Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK

ARTICLE INFO

Article history:
Received 6 July 2010
Accepted 26 July 2010

Keywords:
Natural computing
Soft computing
Machine learning
System
Identification
Condition monitoring
Structural health monitoring

ABSTRACT

A great many computational algorithms developed over the past half-century have been motivated or suggested by biological systems or processes, the most well-known being the artificial neural networks. These algorithms are commonly grouped together under the terms *soft* or *natural* computing. A property shared by most natural computing algorithms is that they allow exploration of, or learning from, data. This property has proved extremely valuable in the solution of many diverse problems in science and engineering. The current paper is intended as a tutorial overview of the basic theory of some of the most common methods of natural computing as they are applied in the context of mechanical systems research. The application of some of the main algorithms is illustrated using case studies. The paper also attempts to give some indication as to which of the algorithms emerging now from the machine learning community are likely to be important for mechanical systems research in the future.

Contents

Well-recognized publisher

A highly-ranked journal, one of the best in the field: IF = 6.4, (2022)

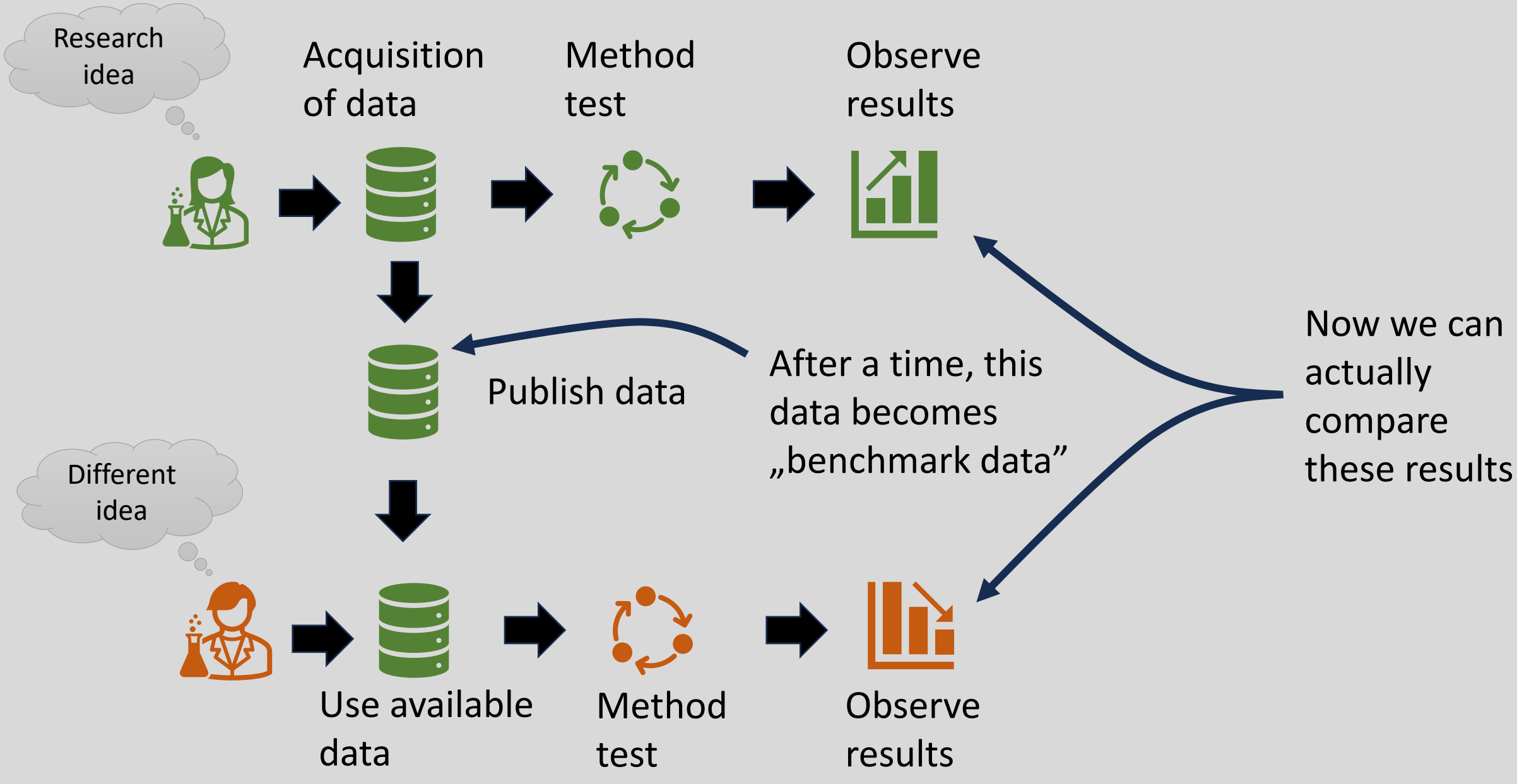
First author is well-recognized: 14500 citations, h-index 54, (2022)

The article is highly cited (200 citations)

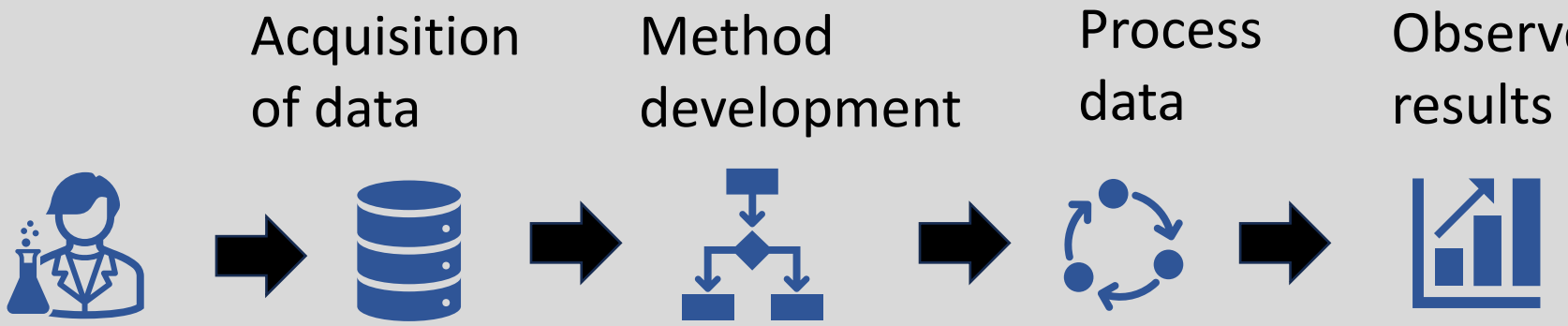
Features of a good paper

- Good English *(not any more… right now everyone is using ChatGPT…)*
- Numerous references including various authors (small self-citation number)
- Reproductible results (Complete information necessary to repeat the experiment)
- Statistical approach (Is the method deterministic?)
- Evaluation of repeatability of results and their transferability to other problems
- High-quality artwork, well-prepared plots
- Benchmark comparison (if possible)
- Blind test (if possible)

Benchmark comparison



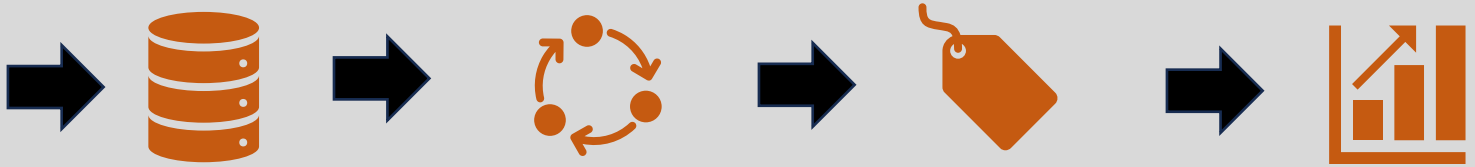
Blind test



Now use new data

Process data

Observe results



Now use new, yet unlabeled data

Process data

Get data labels

Observe results

This is a **blind test**

This is a **double-blind test**

What issues can we encounter in bad papers?

Confirmation bias

- Use of a non-deterministic method „until it succeeds”, report only this final, successful try – *Can be recognized due to lack of statistics or „weird” percentage results*

- Use of a particularily easy example of a problem – for which majority of methods work *– Can be recognized based on lack of comparison with benchmark data or other methods and lack of display of raw data.*

- Use of a full „invention” of a data *– Very rare case, sometimes detectable using source data and using knowledge and experience of other methods in similar category.* ***Deemed a serious breach of scientific conduct***

What issues can we encounter in bad papers?

Lack of proper verification of a method

- One example (a case-study) is used to infer general capabilities of the method – *A very frequent problem. Can easily be detected after analyses of experimental procedures.*

- Lack of false-positive check. E.g. evaluation of a damage-detection method based only on damaged data – *Again, easily detectable based on experimental description*

- Lack of a proper comparison with a benchmark problems and methods or lowering efficiency of a benchmark solutions by their poor configuration – *Lack of benchmark is easily detectable. Lowering efficiency of a compared solution is easily concealable.*

What issues can we encounter in bad papers?

Lack of a blind test (overfitting)

- The method is configured until it obtains the expected result – *Detectable by assessment of metaparameter-setup procedure. If this procedure is either not explained or denoted as a „trial and error approach" – it is possible that overfitting is an issue*

- Testing data are (almost) copies of training data – *Hard to detect without access to raw data. A red flag should be raised every time when authors mention „augmentation of datasets" or „introduction of noise to data" in order to increase data amount.*

Are these issues common?

Can a high-quality journal ensure lack of research errors in its papers?

No! *– it just decreases their probability*

Therefore: *one source is never actually a source. We always need to confirm hypothesis in many papers*

This is why review papers and meta-analyses are so important!

Lets read a paper: title page

Mechanical Systems and Signal Processing 185 (2023) 109823

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Check for updates

Novelty detection approach for the monitoring of structural vibrations using vision-based mean frequency maps

Jakub Spytek *, Adam Machynia, Kajetan Dziedziech, Ziemowit Dworakowski, Krzysztof Holak

Department of Robotics and Mechatronics, AGH University of Science and Technology, Krakow, Poland

ARTICLE INFO

Communicated by J. Baqersad

Keywords:
Novelty detection
Structural health monitoring
Optical flow
Structural vibration
Optimal baseline selection

ABSTRACT

Damage detection is an important part of modern-day engineering. Early damage detection is facilitated by Structural Health Monitoring methods, which may be employed using numerous modalities, such as vibration, guided waves, thermography, or computer vision. These methods produce information that can then be interpreted to detect and localize damage or quantify its extent. Novelty Detection (ND) is a data interpretation approach that enables damage detection without prior knowledge of damage-related influences on gathered data. ND can be conveniently performed using computer vision methods, which allow continuous, non-contact monitoring of large structures with the possibility of relying on such quantifiers as deflection, vibration, or strains.

In this work, we present an ND method for monitoring the structural changes in rotary machinery equipment using vision-based data. The proposed technique detects changes in the characteristic frequencies of vibrations due to damage. Using the optical flow calculated for

Title is always informative

Order of authors is important. Authors are ordered based on decreased contribution, the last one is usually a supervisor of the project

* Denotes a „corresponding author” (someone that knows the most about the method)

Lets read a paper: abstract

Abstract is like a micro-paper that summarizes the most important findings of the paper and provides explanation for why this stuff is important

ABSTRACT

Damage detection is an important part of modern-day engineering. Early damage detectio facilitated by Structural Health Monitoring methods, which may be employed using nume modalities, such as vibration, guided waves, thermography, or computer vision. These meth produce information that can then be interpreted to detect and localize damage or quantify its extent. Novelty Detection (ND) is a data interpretation approach that enables damage detection without prior knowledge of damage-related influences on gathered data. ND can be conveniently performed using computer vision methods, which allow continuous, non-contact monitoring of large structures with the possibility of relying on such quantifiers as deflection, vibration, or strains.

In this work, we present an ND method for monitoring the structural changes in rotary machinery equipment using vision-based data. The proposed technique detects changes in the characteristic frequencies of vibrations due to damage. Using the optical flow calculated for the videos acquired using a high-speed camera, the maps of mean frequency can be estimated and used for evaluating the differences between the reference data set and the data obtained during the monitoring. The Optimal Baseline Selection is used to compensate for the varying operational conditions under which the structure is monitored. The approach was tested on the air compressor working under variable pressure, and the damage introduced to the structure was successfully detected.

Why we want to do something? (motivation)

What do we do?

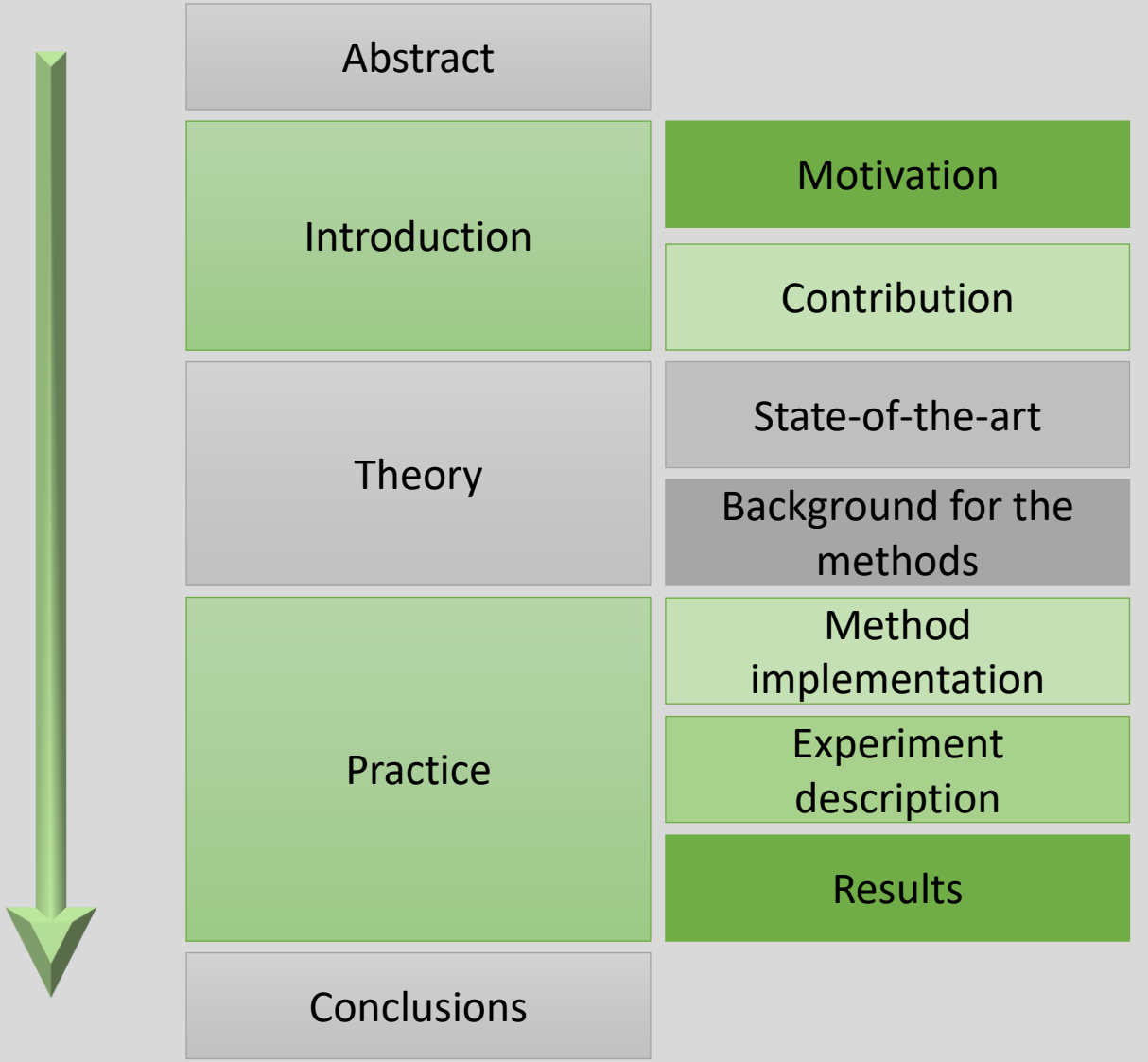
How we do it?

How we test it?

What's the result?

Lets read a paper: paper structure

All of the papers (apart from reviews) follow this structure



Lets read a paper: Steps in order

What you do	What you get
1. Read title, skim through abstract	- Confirmation of topics and relevance
2. Skim through paper. Look for: - artwork quality - section contents - number of experiments - volume of results	- Estimate on **paper quality** and contents - **Tags** for your reference manager
3. Look to „contribution” and „conclusions” sections	- What the authors say this paper is for - What were the most important findings
4. Specialized actions in order to fulfill your current goal	See next slide

Lets read a paper: Steps in order

If you want to broaden your search – look for references that the authors say are relevant. Maybe you will find something new

If you want to understand why this research is important, read the motivation

If you want to understand method capabilities and limitations, look for experimental data and description of results

If you want to repeat the study, look for details of method configuration

If you want to understand the methods' background, look for theoretical section

To sum up: You almost never read a paper page-by-page, from first to last page!

Where to look for papers?

Through AGH library!

We follow this link:
https://bg.agh.edu.pl/zasoby/e-zasoby/e-zasoby-alfabetycznie

And now, if we are at AGH (or use VPN), we can access most of the sources in a full-text mode. We first need to use a particular database/search engine though. I recommend the following:

- ScienceDirect (Elsevier)
- SpringerLink (Springer)
- Wiley

Katalog | Multiwyszukiwarka EDS | Lista A-Z | Wyszukiwarka podstron BG

Znajdź w katalogu online:

Wspólny katalog Biblioteki Głównej oraz bibliotek wydziałowych AGH

Szukaj

Biblioteka Główna Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie
e-mail: bgagh@bg.agh.edu.pl

- Godziny otwarcia
- Sprawdź konto biblioteczne
- E-zasoby
- BaDAP Baza Danych o Autorach i Publikacjach
- RODBUK Repozytorium Otwartych Danych Badawczych Uczelni Krakowskich
- Repozytorium AGH
- Punkt Informacji Normalizacyjnej
- PATLIB Ośrodek Informacji Patentowej
- Portal Historia AGH

O

P

R

S

Science	nauki przyrodnicze, inżynieryjne i ścisłe
ScienceDirect (Elsevier)	nauki matematyczno-przyrodnicze, medyczne, społeczne i humanistyczne
Scientific.Net	czasopisma z zakresu inżynierii materiałowej, materiałoznawstwa, mechaniki
SciVal	narzędzie bibliometryczne do analizy danych z bazy Scopus
Scopus	baza interdyscyplinarna zawierająca wskaźniki bibliometryczne
SIGŻ	rolnictwo, przemysł spożywczy
Social Science Research Network	nauki społeczne (m.in. ekonomia, finanse, marketing, zarządzanie, antropologia, archeologia)
SPIE Digital Library	optyka, fotonika
SpringerLink	medycyna i nauki biomedyczne, nauki przyrodnicze, nauki inżynieryjne, informatyka, chemia, ochrona środowiska oraz nauki społeczne i humanistyczne

T

Search for peer-reviewed journal articles and book chapters (including open access content)

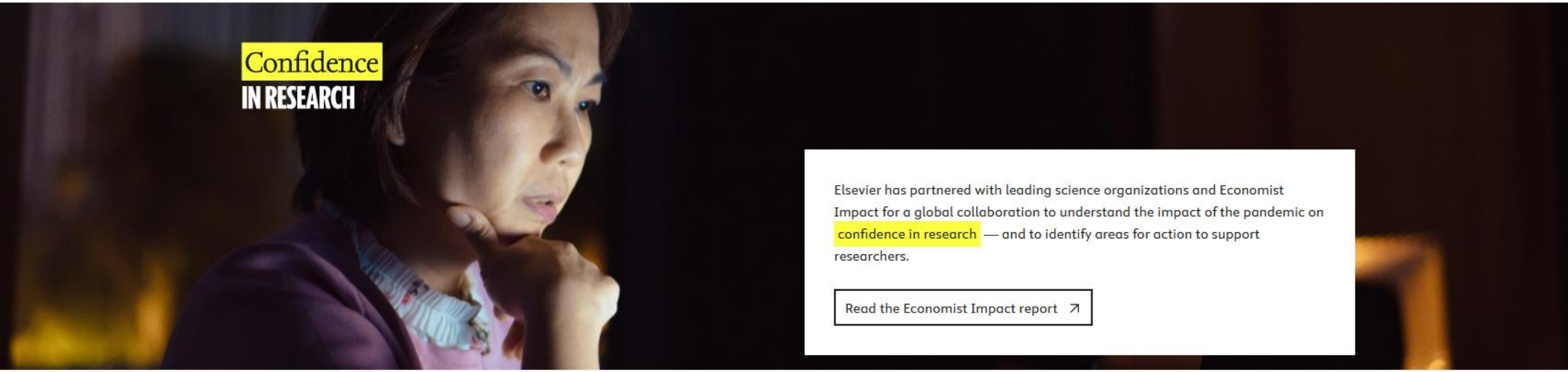
Find articles with these terms

In this journal or book title

Author(s)

Search

Advanced search



Elsevier has partnered with leading science organizations and Economist Impact for a global collaboration to understand the impact of the pandemic on confidence in research — and to identify areas for action to support researchers.

Read the Economist Impact report

Explore scientific, technical, and medical research on ScienceDirect

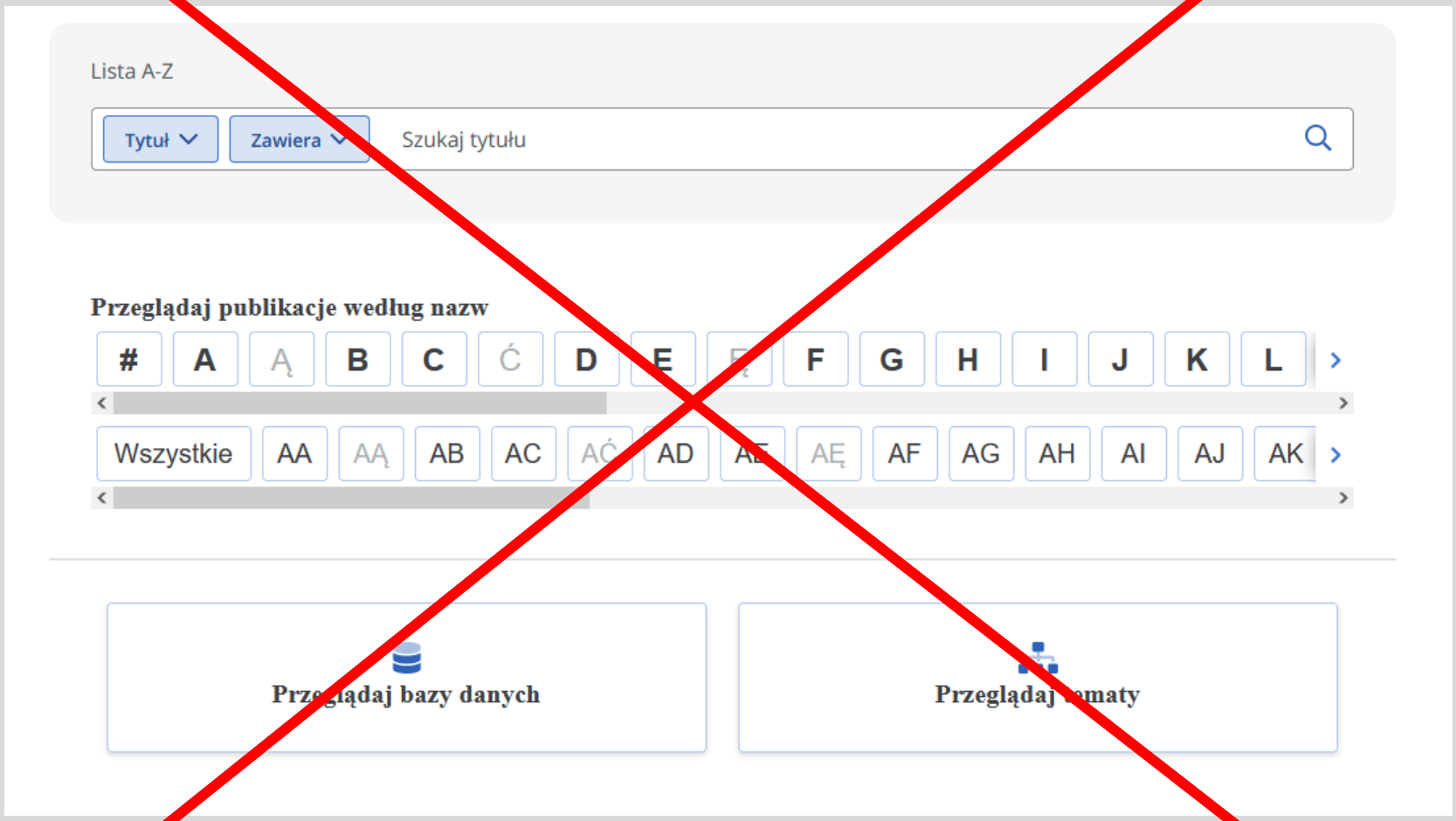
Physical Sciences and Engineering | Life Sciences | Health Sciences | Social Sciences and Humanities

Physical Sciences and Engineering

Chemical Engineering

Chemistry

From foundational science to new and novel research, discover our large collection of Physical Sciences and Engineering publications, covering a range of disciplines, from the theoretical to the applied



I don’t think you want to use „Publication Finder” – It appears that although it should grant you access to articles, it only searches for journals or books. The problem might be on my side though…

How to build a good query

1. Don’t overthink, just use several keywords together, e.g.:

Neural network, damage detection, bridge

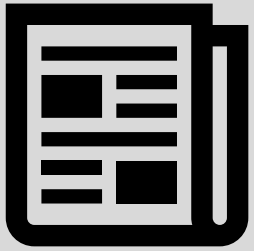
2. Don’t be afraid of hundreds of thousands of results. The best and most relevant ones are usually on top

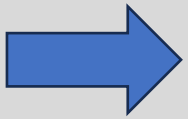
3. Experiment with different wording (it is better to use 15 queries than to craft a perfect wording but only use one query). Store your queries somewhere. They might be worth evaluation and repetition after you learn the area better.

4. Don’t use words that are not likely to be placed in a title or text. „**Good quality**” or „**highly cited**” or „**respected author**” are not good queries.

5. At the beginning of search it is often good to add **Review** or **Survey** – to look for metaanalyses. Later also check for **Benchmark** or **Comparison** – to look for research with a clear reference.

Why should you tag articles?





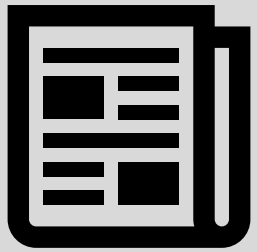
Good quality case-study article on AI-based system for bridge monitoring using vision data, tested in laboratory conditions

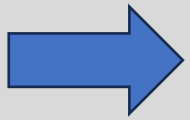
How to store it?
In a folder?
In a subfolder?

Instead, we will use these as tags. This will allow:

- Easy search of articles to support particular claims
- Easy way of making quantitative reviews
- Scalability of a knowledge base
- Easy return to work after long pauses

How should you tag articles?



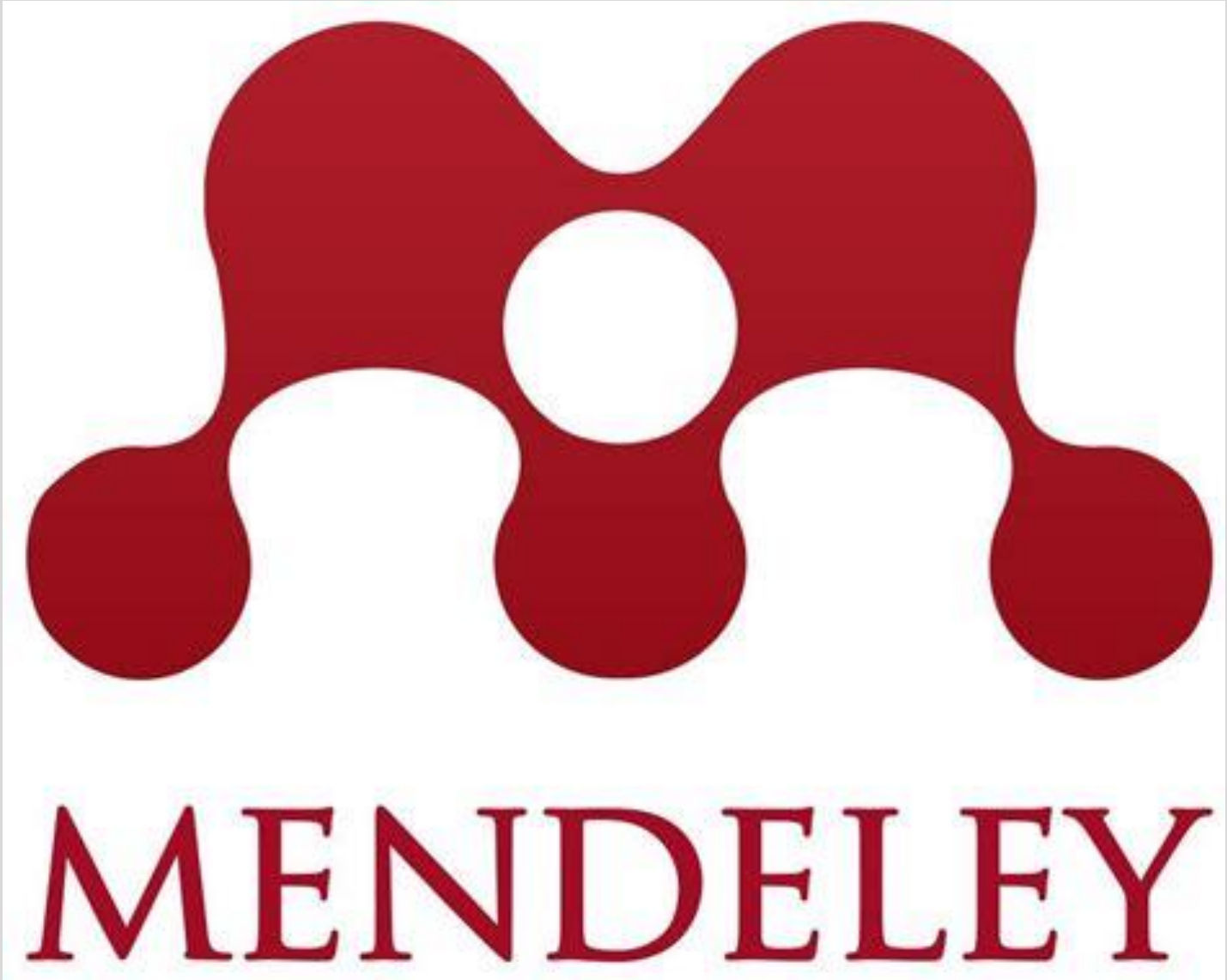


Quality	Category	Main topics	Data type or source	Methods	Application
★★★					
e.g. „High IF" „Conference" „Poor explanation"	e.g. „Review" „Case study" „Proof of concept"	e.g. „damage detection" „vision systems"	e.g. „simulation" „benchmark" „real-life data"	e.g. „deep learning" „fourier-based"	e.g. „bridge"

These should probably be used always

These are my subjective choice

An example…



How NOT to write a review

Let's assume, you decide to spend roughly 35 hours to prepare a literature review for your thesis:

around 1h to 4h:

Database search using a bunch of keywords, skim through abstracts and text
*(download ~**10-20** documents)*

Roughly 1h (often more):

Pick one article and read it.
(Get lost quickly in the chaos of new names and details)

Roughly. 0.5h/article

If the article fits the topic, write down the most important information to a newly developed review
(effectively: write paper's abstract in your own words)

Go back to another article, repeat until all the articles are reviewed
*(rougly **30h** hours in total)*

Result: A chaotic review based on 20 sources that does not have synthesis of any sort and is basically a repetition of abstract information. Spending 2x more time does not affect the result much

How to actually write a review?

a.k.a. : *„Do I really need to read all of these articles?”*

1 – 2 hours:

Database search using a bunch of keywords
*(download ~ **30-50** papers based on titles)*

5 min/paper, 3h in total:

Skim through the papers, Quality and theme assessment, tagging
*(Pick ~**5** „major” papers and ~**10** „supplementary” ones)*

2h/paper, 10h in total:

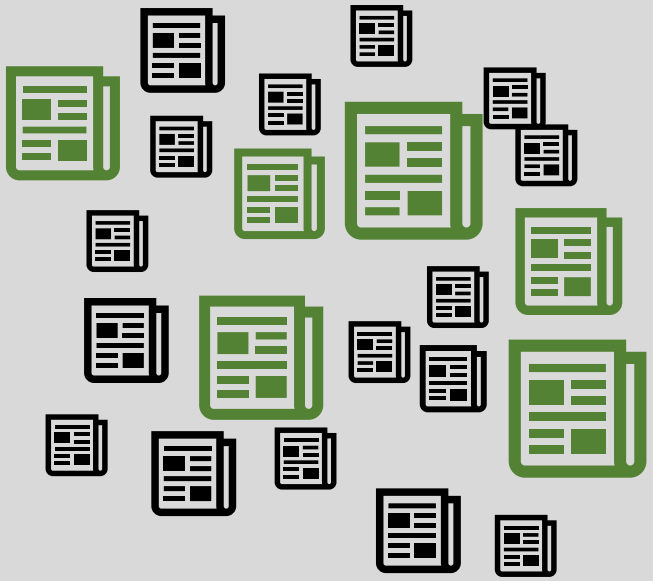
Read through „major” papers
- Mark the most important references
- Check for repeated observations
- Get the main idea (no details)
- What is deemed the most important?
*(Download ~**10** new papers based on refs)*

30min/paper, 10h in total:

Build a scheme of a review, gradually fill it with text
(„major” articles as a cornerstone, „supplementary” ones as examples and additional observations, tags to guide you through your base)

Result: Knowledge synthesis from 25 sources, with self-gained conclusions, using 25h
easy for further development

Features of a good literature review:



- Around 15 – 30 sources (for engineering thesis)
- Reports review work done in the area for further reference
- Knowledge synthesis (added value for the reader)
  (you should probably refer to most of the papers more than once. Avoid simple „combination of abstracts”)
- Critical assessment of knowledge
  „Paper A claims that … but papers B and C say otherwise”
  „Paper D says it is possible, but uses only simulation as a proof”
- Quantitative summaries
  „Many papers use method A [2,3,6,8], but method B is also used for the same purpose [4,9]. There are also rare reports of using method C for this task [5]”

How to work with scientific literature?

How to prepare literature reviews?

How to learn from literature sources?

Ziemowit Dworakowski

AGH University of Krakow