How to conduct a literature survey and how to read a paper – a tutorial

Kai Arras

Social Robotics Lab

University of Freiburg, Germany

April 2015

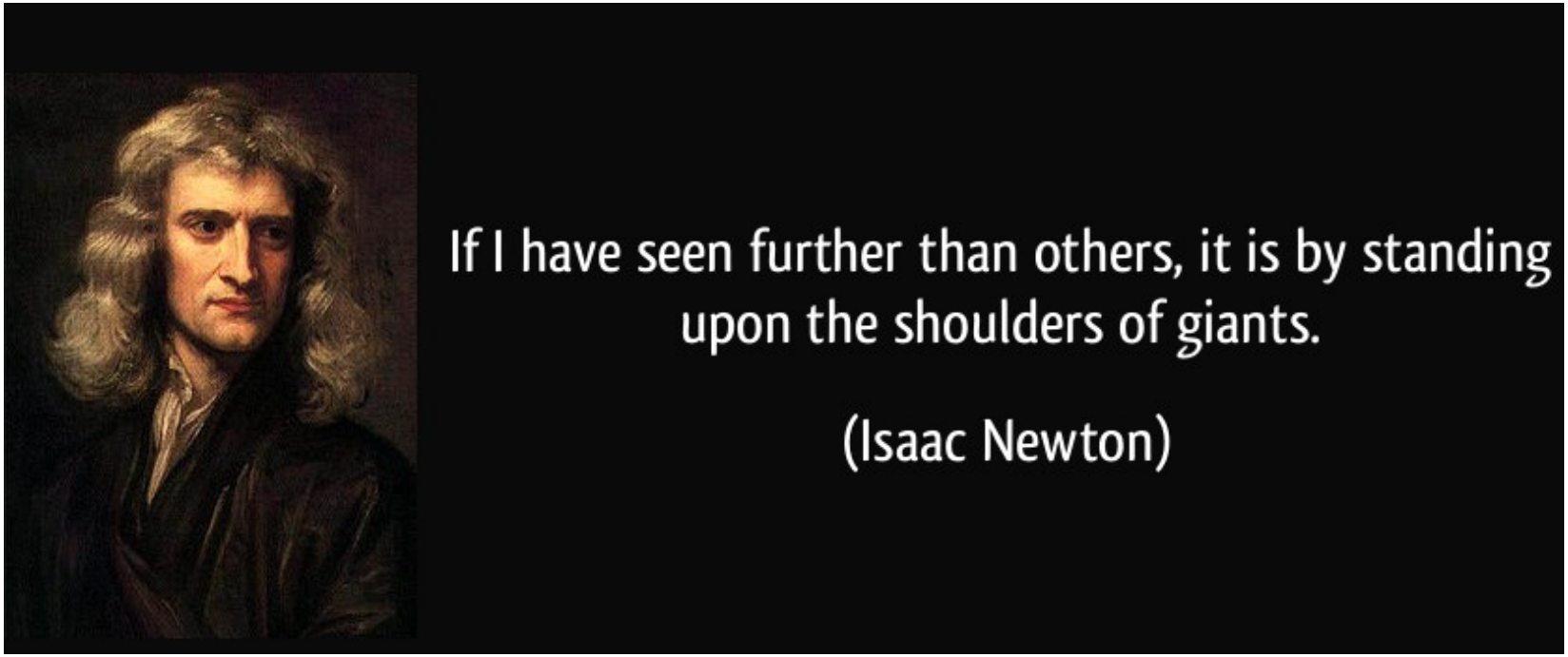
Why This Tutorial?

- Researchers and students spend a great deal of time reading scientific papers. However, this skill is **rarely taught**.
  - This tutorial proposes a practical three-pass method for reading papers in a top-down way at increased levels of detail
- Literature surveys provides a background to a study and put it into **context**. This is key when convincing people of the your ideas
  - This tutorial proposes an procedure for conducting a literature survey that helps to structure and organize your survey and prevent you from drowning in the amount of information
- Largely based on S. Keshav's article, „How to Read a Paper". The proposed methods are extended and combined with techniques from other sources. See references at the end.

How to Conduct a Literature Survey

Why a Literature Review?

- Most ideas don't materialize out of nowhere. They build upon theories, methods, or findings found or developed in **previous work**



How to Conduct a Literature Survey

Why a Literature Review?

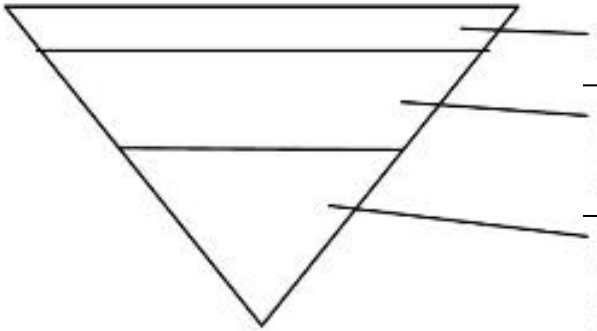
- A literature review is a selection and analysis of existing research which is relevant to your topic
- It shows how your work **relates** to previous research
- You demonstrate the ability to understand and **critically analyze** the background research
- ... and to select and source the information that is necessary to develop a **context** for your work. Concretely:
  - Am I the first one to have this idea? Which aspect of my discovery is new and which one is not? Did someone else have a similar idea? What are the results and conclusions from other similar ideas? Etc.
- Only a literature survey allows you to convince people of the **contribution** of your new idea/finding/theory/method to the field
  - E.g. solve a problem that was unsolved before, fill a gap, generalize an existing theory

How to Structure a Literature Review

- A literature review is not a straightforward summary of everything you have read on the topic
- Neither is it a chronological description of what was discovered in your field
- A common way to approach a literature review is to **start out broad and then become more specific** (see next)
- To deal with a large body of related work, **group similar works** together according to e.g. commonly used theories, methods, system properties, etc.
- This will help to compare and contrast their approaches and facilitates the discussion with the literature in view of your new idea
- Be **careful** and **exact** in dismissing related work when promoting your idea

How to Structure a Literature Review

- A common way to approach a literature review is to **start out broad and then become more specific**



	Example
Broad issues	Robot motion planning
Studies which overlap with your research	Planning in dynamic environments
Studies that are directly related to your investigation	Human-aware robot motion planning

1. First briefly explain the broad issues related to your work. You don't need to write much about this, just demonstrate that you are aware of the breadth of your subject
2. Then narrow your focus to deal with the papers that overlap with your research
3. Finally, concentrate on the research which is closely related to your specific work. Proportionally you spend most time discussing those papers

How to Conduct a Literature Survey

Before you Start

- Identify what you **need to know:**
  - What research has already been done on my topic?
  - What are the relevant subareas of my topic?
  - What are the key papers on my topic?
  - What are the key issues, types of research questions and common approaches?
  - What other research (perhaps not directly on the topic) or other research communities might be relevant?
- Moreover
  - Which research community is relevant for my topic?
  - What are the relevant places for publication (top conferences and journals) in this community?
  - Who are the important authors?

Where to Start?

- First, use a publicly available academic search engine such as Google Scholar, Microsoft Academics, or CiteSeer
- They have become increasingly complete and reliable
- There are **subscription-based scientific citation indexing services** such as ISI/Reuters Web of Science. They offer professionally maintained literature databases for all areas of science with advanced search capabilities.
- Every well-equipped university library has licenses and subscriptions to those services. Contact your **librarian**.
- This is the pro-stuff. However, many pros use Google Scholar, too.
- Here, we will focus on a **pragmatic yet powerful approach** to conduct a literature survey using Google Scholar or similar tools

How to Conduct a Literature Survey

Algorithm

1. Go to http://scholar.google.com (for example)

2. Find some well-chosen keywords to find **five recent papers** in the area („seed“ papers)
   - Select the five papers according to abstract and keyword match
   - Prefer journal articles, then conference papers. Skip text books
   - While you read abstracts and titles, refine your keywords and re-iterate

3. Read the **related work section** of each paper
   - You will find a summary of the recent work in your area
   - If you are lucky you find pointers to a recent survey paper ☺

Algorithm

4. Find **shared citations**, **repeated author names** and publication **places** in the bibliographies
   - These are likely the important papers, authors and places in that area
   - Maintain a statistics of their occurrences
   - Download those papers and set them aside

5. Go to the **website** or Google scholar profile of the **key researchers** and see what and where they have published **recently**
   - That will help you identify the top journals and conferences in that field because the best researchers usually publish in the top places
   - This will also give you more recent high-quality related work
   - Download and set aside the most relevant papers

Algorithm

6. Go to the websites of those journals and conferences and look through their **recent issues/proceedings**
   - This will give you more recent high-quality related work
   - Download and set aside the most relevant papers

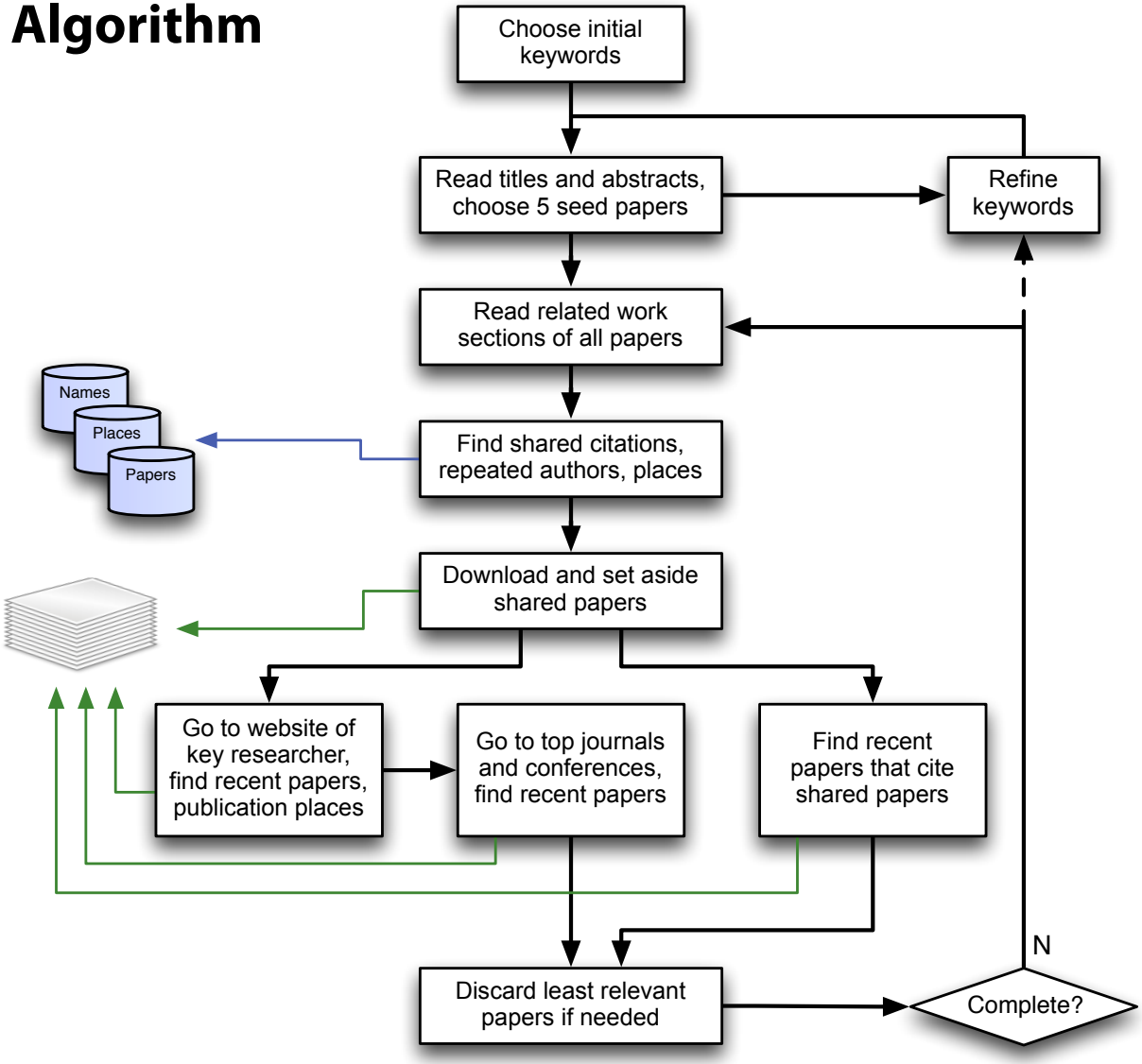
7. Use Google scholar to find recent **articles that cite** the important papers found in step 4 (click on the citation count numbers)
   - Download and set aside the most relevant papers

8. All papers that you have set aside constitute the **first version of your survey**

9. Go to step 3, re-iterate as necessary
   - A good criterion to stop is when the statistics on important author names, publication places and papers has roughly converged

How to Conduct a Literature Survey

Algorithm



After 2, 3 or more iterations you will be able to give robust answers to the questions on slide nr. 7:

- Important papers? authors, publication places? Subareas? Common approaches, main research issues? etc.

The Number of Citations of a Paper

- The number of citations is an important measure to quantify the **importance** of a paper
- The accumulated number of citations of an author measures the **impact** of his/her work
- However, the number is not robust against **one-hit wonders**, does not measure durable academic performance
- Hirsch's h-index aims to provide a **robust** single-number metric of a researcher's impact
- Note that numbers of citations and h-indices scale with the **size** of a community. In large communities, it is easy to receive many citations
- Rules of thumb: papers with >1000 citations are **seminal** papers, papers with >100 citations are **important** papers. Of course, this depends strongly on the paper's age

The Number of Citations of a Paper

- When you conduct a literature survey, the number of citations of a paper or an author is a **good heuristic** to find important papers and authors
- However:
  - Very recent papers had no time to accumulate many citations
  - Young (but promising) researchers have rarely many citations
  - Sometimes, a great paper is not cited because it is ahead of its time, proposes an uncommon idea against the current mainstream, or other reasons of **human error**

... Once the Survey is Done:

- Reading and understanding a scientific paper is **not easy**. In most cases you do not have all the background knowledge required to understand a paper
- Understanding a paper is not a **yes/no condition**: you must decide when you have a reasonably good understanding of the content
- Apply a **top-down approach** when reading the paper. Try to get an overview and then focus on the details (see next)
- Papers might also contain mistakes
- Read a paper critically. **Analyze** rather than describe
  - **Descriptive**: summarizing other researchers' work without saying what their findings mean (to the field or your topic)
  - **Analytical**: judging relative merits of other researchers' work, discussing limitations. Recognizing the possibility of taking research further

Before you Start

- Again, identify what you **need to know:**
  - What is the **research problem** addressed in the paper?
  - Why is this problem important?
  - What are the **novel ideas** and key concepts proposed by the authors?
  - How does the paper relate to **other** papers?
  - **What is the main contribution of the paper?**
  - Are there concepts or methods in the paper that are **unclear**?
  - What are the **flaws** or limitations of this paper?
  - How would you **extend** this work?

An Efficient Three-Pass Method

- The key idea is that you should read the paper in up to **three passes**, instead of starting at the beginning and plowing your way to the end

- Each pass accomplishes specific goals and builds upon the previous pass:

    - The **first pass** gives you a general idea about the paper
    - The **second pass** lets you grasp the paper's content, but not its details.
    - The **third pass** helps you understand the paper in depth

An Efficient Three-Pass Method: First Pass

- The **first pass** is a quick scan to get a bird's-eye view of the paper
- You can also decide whether you need to do any more passes or to not read further
- This pass should take about **five** to **ten minutes** and consists of the following steps:
    1. Carefully read the title, abstract, and introduction
    2. Read the section and sub-section headings, but ignore everything else
    3. Read the conclusions
    4. Glance over the bibliography, mentally ticking off the papers you've already read

An Efficient Three-Pass Method: First Pass

- At the end of the first pass, you should be able to answer the five *C*'s:
  - **Contribution**: What are the paper's main contributions?
  - **Category**: What type of paper is this? A paper that proposes a new method or theory, a survey paper, a data set paper, a description of a research prototype?
  - **Context**: Which other papers is it related to? Which theoretical fundamentals were used to address the problem?
  - **Correctness**: Do the assumptions appear to be valid (and relevant for me)?
  - **Clarity**: Is the paper well written?
- Using this information, you may choose not to read further. This could be because the paper doesn't interest you, makes invalid assumptions or you don't know enough about the area to understand the paper

An Efficient Three-Pass Method: Second Pass

- In the **second pass**, **read the paper** with greater care, but ignore details such as mathematical derivations or proofs
- It should take up to **an hour** and help to jot down the key points, or to make comments in the margins, as you read
  - Look also carefully at the figures, diagrams and other illustrations. Are they easy to understand and meaningful? Sloppy graphs, missing error bars, or unclear illustrations of results will separate rushed, shoddy work from the truly excellent
  - Remember to mark relevant unread references for further reading – a good way to learn more about the background of the paper
- After this pass, you should be able to grasp the content and main findings of the paper
- This level of detail is appropriate for a paper in which you are interested, but does not lie in your exact research speciality

An Efficient Three-Pass Method: Second Pass

- Sometimes you won't understand a paper even at the end of the second pass
- This may be because the subject matter is new to you, with **unfamiliar terminology** and acronyms. Or the authors may use a theory, method or experimental technique that you don't understand
- The paper could also be **poorly written** with unsubstantiated assertions or poor writing style, structure or language
- Or could it just be that it's late at night and you're **tired**?
- You can now choose to
  - a) **ignore** the paper, hoping you don't need to understand the material to be successful in your career
  - b) return to the paper **later**, perhaps after reading background material
  - c) persevere and go on to the **third pass**

An Efficient Three-Pass Method: Third Pass

- To fully understand a paper, you require a **third pass** in which you carefully read the entire paper
  - This pass requires great attention to **detail**. You should identify and challenge every assumption in every statement
  - During this pass, you should also jot down ideas for **future work**
- At the end of this pass, you should be able to reconstruct the entire paper from memory, and – if the paper is complete and well written – be able to **reimplement** the proposed method (for CS-related papers)
- In particular, you should be able to identify its **strengths** and **weaknesses**, pinpoint implicit assumptions, missing citations to relevant work, and potential issues with experimental or theoretical techniques
- Can take up to **5 hours** or more for beginners, about 1-2 hours for experts (e.g. for short/conference papers)

An Efficient Three-Pass Method

- In summary, the three-pass method...
  - Enables students to efficiently and effectively answer the key questions on main contribution, problem addressed, motivation, methods used, weaknesses, etc. (see slide nr. 16)
  - Prevents students from **drowning in the details** before getting a bird's-eye view
  - Allows to **estimate the amount of time** required to review or read a set of papers
  - Allows to adjust the **depth of paper evaluation** depending on personal needs and how much time is available

References

Sources

- S. Keshav, **How to Read a Paper**, ACM SIGCOMM Computer Communication Review 37(3), p. 83-84, 2007
- K. Haq, **Managing and Reviewing the Literature**, Lecture Notes Research Skills Workshops, Graduate Research School, The University of Western Australia, 2014
- **Starting a literature review**, University of Reading, Study Advice and Maths Support, URL: https://www.reading.ac.uk/internal/studyadvice/StudyResources/Essays/sta-startinglitreview.aspx, last visited April 2015