RESEARCH METHODS

An Introduction to Research and Scientific Paper Reading and Writing

Diego Calvanese
KRDB Research Centre for Knowledge and Data
Free University of Bozen-Bolzano

Why are you attending this course?

- Get an understanding of what research is
- Learn about the methods used in research
- Learn best practices and get useful advice
- Get inspiration for our PhD and for our future career
- Participate actively, and learn through practice

Structure of the course

1. Good scientific writing style [E. Franconi]
   27/10/2014 - 31/10/2014
2. Introduction to research and paper writing [D. Calvanese]
   3/11/2014 - 7/11/2014
3. Empirical/experimental CS research methodology [F. Ricci]
   10/11/2014 - 14/11/2014
4. Writing a research plan / proposal [S. Helmer]
   17/11/2014 - 21/11/2014
5. Research evaluation and reviewing, bibliometrics [W. Nutt]
   24/11/2014 - 28/11/2014
6. Presenting scientific work [M. Montali]
   1/12/2014 - 5/12/2014

Credits

- Paul Wagner, University of Wisconsin
- Prasant Mohapatra, University of California, Davis
- Simon Peyton Jones, Microsoft Research Cambridge
- Gordana Dodic-Crnkovic, Mälardalen University, Sweden

Disclaimers

- No one-size-fits-all solution
- No recipe for success
- Success or failure is often difficult to define and assess
- Some advices may not fit individual needs
- There is no black and white
- We cannot teach how to get excited and passionate

  … but we can provide examples
- Quest for learning may trade-off other activities

What is research?

Definition by Merriam-Webster:

1. careful or diligent search
2. studious inquiry or examination; *especially*: investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or laws in the light of new facts, or practical application of such new or revised theories or laws
3. the collecting of information about a particular subject

Constituents of research

- Defining, redefining, and/or formalizing problems
- Formulating hypotheses
- Suggesting solutions or solution approaches
- Collecting and analyzing data
- Experimenting
- Eventually validating the hypotheses and/or
- Deducing new conclusions
- Deriving new knowledge and/or formulating new theories

Research is not just coming up with a problem and solving it, but devising methodologies for its solution.

What else is NOT research?

The following is not research per se:

- Playing with technology
- Developing code
- Deploying standard or commercial technology
- Doing what others have already done

- However, each of these can be done as part of research

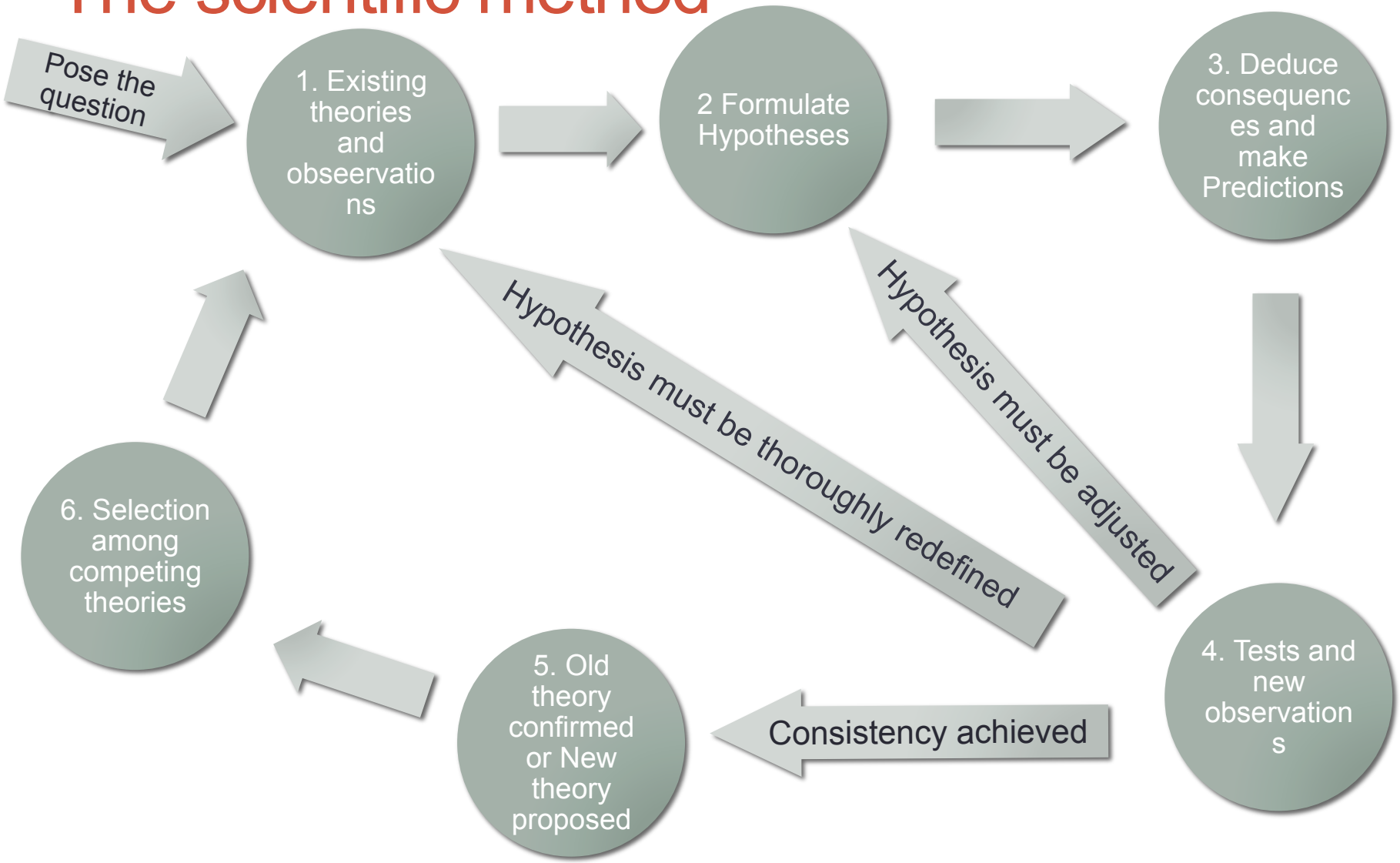
Keywords associated to research

- Culture: Research is a culture, which requires continuous nourishment and practice through quest for innovation.
- Attitude: Requires hard work, dedication, perseverance, and an appropriate attitude.
- Motivation: Research planning cannot really follow a “scheduled” approach.
- Dignity: Not everyone has to do research, and research should not be pursued for glorification.

Why are we doing research?

- Intellectual satisfaction provided by doing something innovative and creative.
- Meaningful and long-lasting contributions towards the advancement of mankind and society.
- Attain a higher level of understanding of fundamental concepts.
- Enjoy the challenges of solving unsolved problems.
- Degrees, financial benefit, and respect all come along the way.

How is research done?
The scientific method



Different types of research - 1

- Descriptive
  - surveys, comparative and correlational methods
- Analytical
  - analyze and critically evaluate information
- Applied
  - address practical problems and solutions that can be implemented for near-term benefits
- Fundamental
  - generalization and formulation of theories

Different types of research - 2

- Quantitative
  - numerical results are used to validate claims
- Qualitative
  - comparative development of usage patterns and experiences
- Conceptual
  - relies on abstract ideas or theories
- Empirical
  - relies on experience and observations

Additional types of research

- Combinations of two or more of the previous types
- Diagnosis or sensitivity studies
- Exploratory
- Decision-oriented
- Laboratory research
- Historical
- Discovery of a unique phenomenon

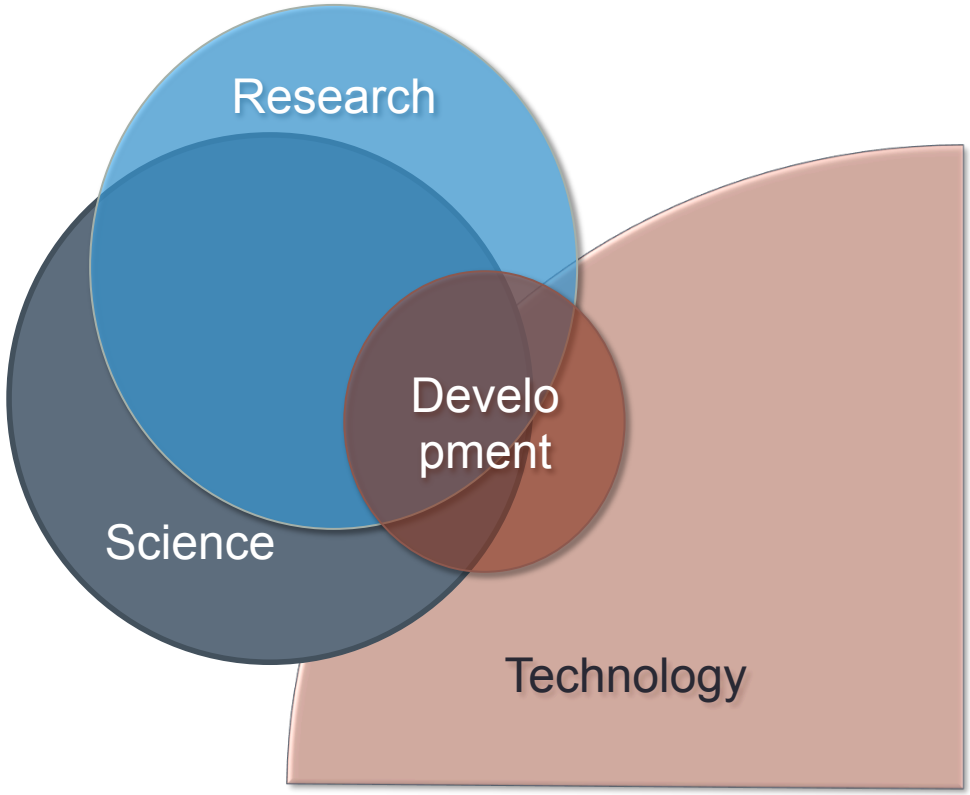
Research in Computer Science

- We want to look into the specific aspects of Computer Science research
- We need first to understand:
  - what CS is
  - what the different branches of CS are …
  - and how they differ

What is computer science?

1. The discipline of Computing is the systematic study of algorithmic processes that describe and transform information: their theory, analysis, design, efficiency, implementation, and application [P.J. Denning, 1989]
2. Computer Science is the study of phenomena related to computers [Newell, Perlis and Simon, 1967]
3. Computer Science is the study of information structures [Wegner, 1968, Curriculum 68]
4. Computer Science is the study and management of complexity [Dijkstra, 1969]
5. Computer Science is the mechanization of abstraction [Aho and Ullman 1992]
6. Computer Science is a field of study that is concerned with theoretical and applied disciplines in the development and use of computers for information storage and processing, mathematics, logic, science, and many other areas [M.S.Mahoney 1992]

Science vs. Technology



[Gordana Dodic-Crnkovic. Scientific Methods in Computer Science. Proc. of the Conference for the Promotion of Research in IT. 2002]

Branches of Computer Science

Very broadly, we distinguish between:

- Theoretical CS
- Empirical (or experimental) CS

Both branches have in common modeling:

- To study a phenomenon, it must be simplified
- Irrelevant features are abstracted away, and only the relevant ones are modeled
- Theoretical ground helps identifying the relevant features
- Using the model, we can predict observable/measurable consequences of given changes

Crucial questions in modeling

- How to model?
- What to take into account? What to neglect?
- What formalism/language to use in modeling?
- Does the model serve its purpose?
- Do we have the right level of abstraction/resolution?
- How does behavior of the model differ from expectation?
- How does the model differ from “reality”?
- Are the results obtained through the model valid?
- How does a new model behave wrt an old one?

Theoretical Computer Science

- Adheres to the traditions of logic and mathematics
- Builds theories as logical systems with the aim of deriving/proving theorems
- Relies on models and levels of abstraction
- CS theories do not compete with each other to explain the nature of information
- New theories do not aim at reconciling theory with experimental results
- The basic model of computation is not questioned
- Theoretical results are judged by:
  - the insights they reveal about computing models
  - their utility for computing, and/or
  - their ease of application
- Central topics are:
  - design and analysis of algorithms
  - understanding the limits of computation
  - distillation of knowledge acquired through conceptualization, modeling, and analysis

Empirical Computer Science

- The field of inquiry is the nature of information processes
- Experiments used for the following activities:
  - for theory testing and explanation
  - where theory and deductive analysis fail
  - to unearth new phenomena that need explanation
  - to help derive theories from observation
- Experiments are made essentially in every area of CS
- In HCI and software engineering, experiments need to involve also humans

Selection of the PhD topic

- Understand your expertise and limitations
- Focus first on breadth of knowledge; depth comes later
- Be specific about the topic, but be flexible about the scope
- The topic should be rich enough but also accessible
- Do not set rigid barriers regarding the topic
- The topic may not be discipline specific
- Be prepared to revisit the topic you selected, if needed

Expanding knowledge

- You should expand your breadth of knowledge on the selected
- Read the fundamentals on the topic to build up the foundations for your research
- These efforts will pay off in the long run, even if it does not seem so initlially
- Feel open about broadening the scope of the topic as you build up on it

Building foundations

- You should have or build a very strong background and foundations on the broad area of your topic
- You should have some ideas about the state of the art
- You should like and enjoy the chosed topic
- New topics vs. old topics
- You need to envision the future prospects of your intended topic (e.g., working on a standard that later is dropped)

Reading papers

- Researchers must read a lot
  - to cultivate knowledge of the area
  - to learn about recent advances
  - to avoid reinventing the wheel
  - to place your work in the proper context (related work)
- Different papers require different levels of attention
- Careful selection of papers saves a lot of time
  - there are several must-read papers in your area, including “historical” papers
  - there is a lot of “garbage” around

Three-pass approach for reading papers

1. Quick scan (5-10 mins)
   - To decide whether the paper is worth reading at all
   - Reduces significantly the number of papers to process further

2. Reading with greater care (1 hour)
   - Helps in grasping the content
   - Helps to better understand the contributions of the paper

3. Detailed reading (4-5 hours, but may take much longer)
   - To fully understand the paper
   - Helps in identifying open issues and ideas for future work

[S. Keshav, ACM SIGCOMM Computer Comm. Review. 37(3), 2007]

Reading papers: Pass 1

- Carefully read title, abstract, and introduction
- Read section and subsection headings, but ignore everything else
- Read the conclusions
- Glance over the references
- At the end, you should be able to answer the following:
  - Category: what type of paper is it (experimental, system descr., …)
  - Context: to which papers is it related? What bases were used?
  - Correctness
  - Contributions
  - Clarity

Reading papers: Pass 2

- Read with greater care, but ignore details (e.g., proofs)
- Identify areas of your interest
- Identify results relevant to the scope of your topic
- Scribble in the margin important points, thought, questions
- Mark relevant references for further reading

After this pass, you should be able to:

- Grasp the content of the paper
- Summarize main contributions, with supporting evidence
- You might not understand the paper, and the reason might be that it is badly written

Reading papers: Pass 3

- This is required to fully understand the paper, especially if you have to review it
- Try to virtually re-implement the paper
  - Make the same assumptions as the authors
  - Re-create the work, re-prove the results, …
  - Compare the re-creation with the actual paper
- This requires great attention to detail
- Think how you would present a particular idea
- You can learn new techniques, and good presentation style
- Note down open problems and ideas for future work

Paper reading assignment

- Propose two papers:
  1. from the top conference or journal in your research area, possibly with a high number of citations (sign of good quality)
  2. from a less known conference or workshop in your research area

- Apply to both the 3-pass reading approach (answering the respective questions)
  1. for paper 1, apply all three passes
  2. for paper 2, stop after pass 2
- In addition, identify (and note down) also weaknesses of the paper and points where to improve
- Apply pass 1 to the references in each of the papers, and use it to prepare a list of follow-up readings

Where to publish

It is important to select the right venues where to publish

- publishers
  - reputation
  - professional societies
  - stay away from private commercial publishers (predatory publishing!)
- quality and metrics
  - impact factor
  - citations, bibliometrics
  - accessibility to researchers
- electronic publishing
- make your work available online

Citations and impact factors

- Citations quantify the impact of a paper
- The number of citations
- Impact factor (IF) is used mostly for journal, and is defined as the average number of citations for its papers
- Impact factor is sometimes misleading (e.g., predatory publishers force citing papers in their journals)
- More details about bibliometrics will be covered in Werner Nutt's part

Workshops, conferences, journals

In computer science, we often follow the 3 phase model

1. One or more workshop papers with initial ideas and preliminary contributions
2. One or more conference papers, each providing original, substantial results
3. A journal paper, that consolidates, and expands the original research contributions

How to write a good research paper

Seven advices by

[S. Peyton Jones, Microsoft Research Cambridge]

1. Don’t wait, write
2. Identify your key idea
3. Tell a story
4. Nail your contributions
5. Related work comes later
6. Put your readers first (examples)
7. Listen to your readers

http://research.microsoft.com/~simonpj/

Paper writing assignment

- Consider paper 2 that you have selected for your reading assignment
- Rewrite abstract, introduction, and conclusions, aiming at correcting the weaknesses that you had identified
- Try to integrate the list of references