Introduction
SCC5933 – Research Methodology in Computer Science

Prof. Moacir Ponti
www.icmc.usp.br/~moacir

Instituto de Ciências Matemáticas e de Computação – USP

2018/1

Introduction

What is research?

- "**Research** is the process of **gathering information** about some subject, analyze them using the **scientific method** with the intention of **increase the stock of knowledge**" (Wikipedia)

Introduction

Method

- "The **scientific method** is a set of **basic rules** for a scientist to develop a controlled experiment in order to test and observe events, so that to reach conclusions and report those conclusions, that, in case of validity, are then applied to science"

(Wikipédia)

Choose the research theme

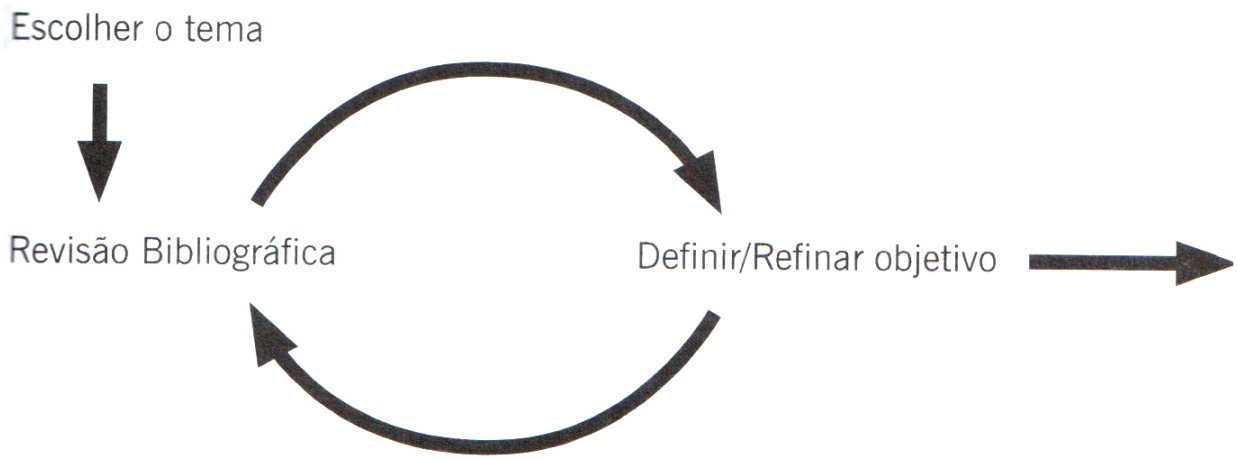
Theme

- The choice can be made looking for:
  - **Relevant**: scientific, social, technologic,
  - **Adequate**: to those employed at university, institute and research lab

- Check for time and feasibility to develop the research
  - **Scope**: it is not necessary to solve it all. It is better to limit then to be too broad.

Objective

- The **objective** can be defined with a **literature review**
  - Should be **an action** that addresses some gap or existing problem
  - Must be coupled with a well defined **hyphotesis**



Objective

- **Warning**: objectives as **proposal** are usually **weak**
  - If the objective of research is "to propose something", then the mere proposal is sufficient?

- Enunciate the problem in a precise way
- Explain why the problem is important given the literature review
- Make sure the premises are sound

Some examples

- "...this project *proposes* the use of optimization methods for vehicle route problems..."
- "...the main objective is to *develop* neural network algorithms for sentiment analysis in text..."
  - It is hard to grasp the actual problem to be addressed
  - It is not clear what exactly is the research question and its importance in those objectives

Hyphotesis

- Good objectives are driven by good **research hypothesis**

Hypothesis

- **Claim** that will be tested to be **true or false**
  - The research project must investigate the claim in order to **confirm** or **falsify** this claim;
  - Defining a sound hyphotesis is what differentiates research from a technical work.

Literature Review

- The research must keep reading throughout the research projects
- It is ok to start with books and surveys
- After you master the main techniques, then search for relevant work on good repositories
- Read papers in a critical way:
  - LARAMEE, R. S. How to Read a Visualization Research Paper: Extracting the Essentials. IEEE Computer Graphics and Applications, Vol. 31, No. 3, 2011, pages 78-82. Disponível em `http://www.cs.swan.ac.uk/~csbob/research/how2read/laramee09how2read.pdf`.
  - FOWLER, M. How to Read Signal Processing Journal & Conference Papers. `http://www.ws.binghamton.edu/fowler/HowReadPapers.htm`.

Repositories

- Scholar (`http://scholar.google.com`)
- Scopus (`http://www.scopus.com`)
- Web of Science (`http://www.webofknowledge.com`)
- ...

Evaluation

How to evaluate your research?

- Define, as soon as possible, how to measure your results in order to understand how close you are to the main objetive
  - Try hard, but if necessary, drop/change the initial idea.

- Since usually $\sim 90\%$ of outcomes are actually failures, we have to make sure we are evaluating correctly the results, since the beginning

- Understand all research has **limitations and weak points**
  - Example: little innovation, incremental contribution, results marginally different from state-of-the-art, application is restricted, it is not scalable, etc.
  - But: a negative result is also **good** if the method was correct!

To be exposed to research ideas

- Exposition makes it easier to (re)define objectives and find ideas
  - Discuss your work with other colleagues and researchers
  - Frequently read papers (at least 1 monthly)
  - It is the responsibility of the student to bring ideas and possibilities to discuss with the supervisor!
  - Follow important researchers on social media (twitter, research gate, etc.)

Agenda

Introduction

Steps of the scientific process

Research levels/tipes in Computer Science

1: “Product or implementation”

- Can be innovative or not
  - If lacking a hyphotesis, then it is not
- When innovative, it is usually exploratory
- If it is a system or reproduction, can be reported in a “Technical Report”
- Acceptable for undergraduate final project (TCC), but hardly for Master or Doctorate degrees

Exploratory

- It is acceptable to not compare with previous work

- Biological computer that solves problems such as the travelling salesman

Vic Norris et al. *Computing with bacterial constituents, cells and populations: from bioputing to bactoputing*. Theory Biosci. 130(3): 211-228, 2011.

System or implementation

- Can be justified when there is a clear application, not yet explored
  - Health,
  - Education,
  - Agronomy,
  - ...
- But, if so, then it is important to compare with previous work at least qualitatively!

2: “Something different”

- proposes a “different approach” to some problem, or a “different” implementation or application
- require literature review and qualitative comparison

- **Can be a different approach**, *not necessarily better*
- It is usual in problems that are well studied, but not sufficiently
- Can result in case studies

Deep Learning

- For a while (~ 4 years), just approaching something with deep learning was enough
- Now those need more rigor

Thanks to `www.xkcd.com`

- It is valid to propose "something different" when there is scarce data or time escassos.
- A well defined case study can be a good way to start
- But, make sure
  - **premises** are convincing
  - there is some hyphotesis

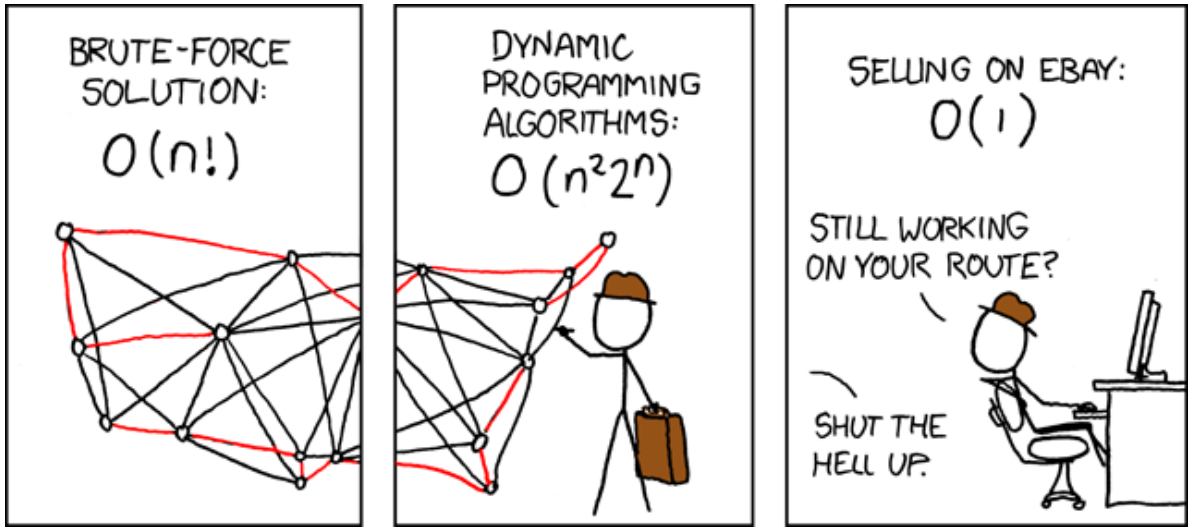
3: "Something supposedly better"

- A problem that is well studied, there is available data and papers reporting results on those well known datasets
- Need to follow protocol, evaluation measures, that are previosly defined by the literature

- When there are already many solutions: you must justify why your approach is valid or better in some sense
- Often results in an incremental contribution.
  - need a better discussion on the drawbacks and advantages

Travelling salesman problem

- Formulated in 1930, NP-hard. Brute-force solution is$O(n!)$



Thanks to: `www.xkcd.com`

Example: image denoising

Noisy image

State-of-the-art

Our method

4: "Something better"

- New results are better considering standard tests
- Datasets that are known and widely used in the literature
- Comparison is direct since everyone follows the same protocol
- Advances the state of the art

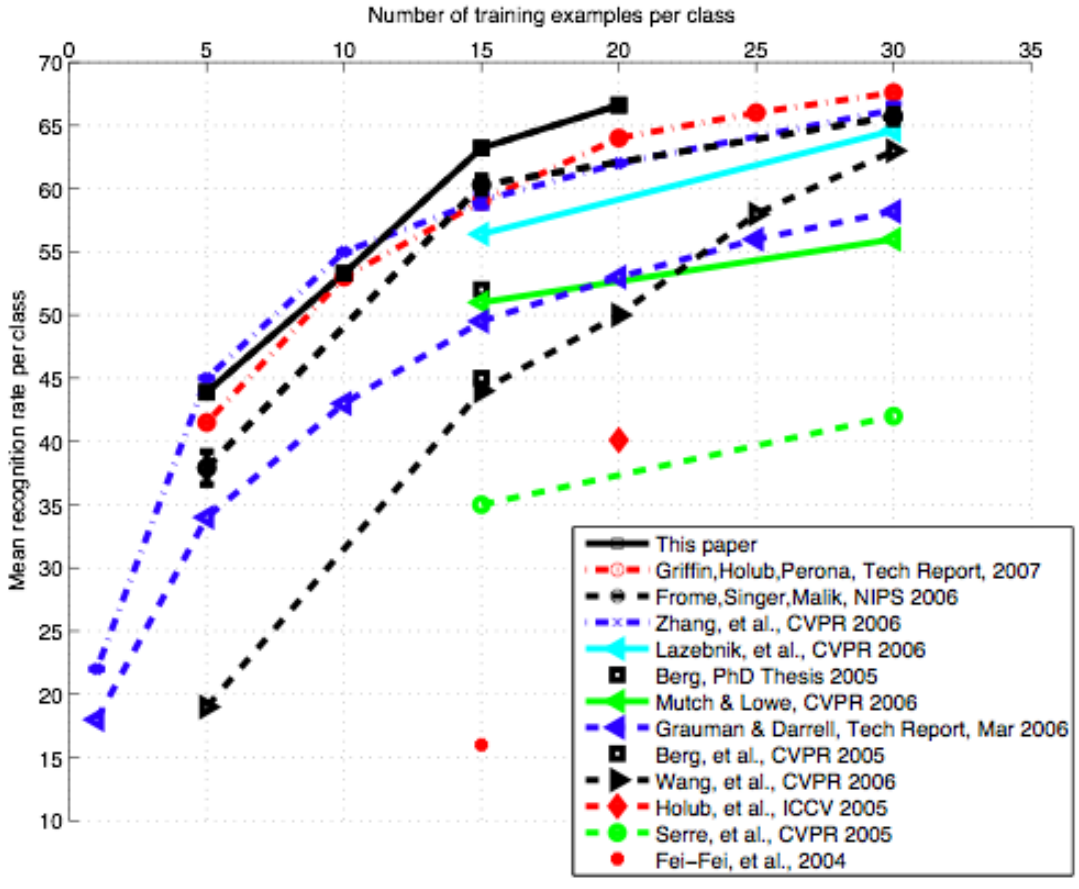
e.g. image classification

- Benchmark: Caltech-101, ImageNet


adidas
JABULANI


Jay fro

- If your method is better, then you add a new 'line' among the most relevant ones



5: “Proof”

- Need a good theoretical background
- Involves the use of theorems, lemmas, in order to address some problem under some premises

- Following some theoretical framework, write a proof based on induction, deduction, contradiction, etc.
- Modern computer science was born with such types of research, in the decades of 1930-1940
- First themes were: computability, algorithms, complexity, information theory, optimization, artificial intelligence, etc.

Alan Turing. On computable numbers, with an application to the Entscheidungsproblem. Proc. London Math. Society, vol. 42, 230–265, 1937

A remarkable example

$\mathbf{P}$ vs. $\mathbf{NP}$

- A problem for which there is an algorithm that **finds a solution** in polynomial time: class $\mathbf{P}$
- A problem for which there is an algorithm that **verifies a solution** in polynomial time: class $\mathbf{NP}$
- Prove if $\mathbf{P} = \mathbf{NP}$ is one of the most relevant open problems in computer science.

Another example

Compilers

- Check machine code for 64 bits and multicore processors
- Source code correctness

Computer Science research

- **Theoretical**,
- **Empirical**,
- **Exploratory**.

The role of the supervisor

- Offer criticism
- Help interpreting and discussing the results, facilitating new ways to solve some problem
- Recommend studies and papers
- Read and give input to text, thesis, dissertation, papers, reports