Advice for a Successful PhD Experience

ICDAR Doctoral Consortium
Kyoto, Japan
November 11, 2017

Professor Daniel Lopresti
Computer Science and Engineering
Lehigh University
Bethlehem, PA , USA

Acknowledgements

Much of this presentation is derived from these sources:

- How To Write A Dissertation – or – Bedtime Reading For People Who Do Not Have Time To Sleep
  http://www.erp.wisc.edu/profdev/how_to_write.pdf
- Dissertations
  http://www.unc.edu/depts/wcweb/handouts/dissertation.html
- How To Write a Good (no, Great) PhD Dissertation
  http://www.cs.cmu.edu/~priya/ICSOC-PhDSymp-2006-dist.pdf
- How to Write Up a Ph.D. Dissertation
  http://www.cs.jhu.edu/~jason/advice/how-to-write-a-thesis.html
- Success in Graduate Student Research
  http://www.cse.lehigh.edu/~lopresti/Resources/GradSuccess.pdf

Disclaimers

- The comments / rules / opinions / philosophies you will hear today reflect my own personal biases.
- Biases are strongly field-specific (e.g., CS vs. biology).
- Biases are often even subfield-specific (e.g., robotics vs. networking vs. bioinformatics vs. AI).
- Biases are also country-specific (e.g., U.S. vs. Japan).
- In the end, your most important guide is your advisor.

Some things I say may sound harsh – this is serious business.

But first …

The International Conference on Document Analysis and Recognition (ICDAR) is the main conference in our field.

ICDAR2017

The 14th International Conference on Document Analysis and Recognition

November 10-15, 2017

Kyoto Terrsa, Kyoto, Japan

You are lucky to be here – take full advantage of the opportunity!

http://http://u-pat.org/ICDAR2017/

What is a PhD?

- A substantial body of original, ground-breaking work (can be in one field or interdisciplinary).
- Proposes a hypothesis and provides arguments to substantiate or refute that hypothesis.
- Evidence you can do research that matters (work that makes a difference in the field, that people care about).
- Should contain 2-3 key ideas you can articulate at the drop of a hat (think "elevator speech").
- A significant piece of independent writing that you want to be proud of for years – or even decades – to come.

From http://www.cs.cmu.edu/~priya/ICSOC-PhDSymp-2006-dist.pdf

What will you learn in the process?

- All scientists need to communicate their discoveries; the PhD dissertation provides intensive training in communicating with other researchers.
- Writing a dissertation requires a student to think deeply, to organize technical discussions, to muster arguments that will convince fellow scientists, and to craft rigorous, formal arguments.

From http://www.erp.wisc.edu/profdev/how_to_write.pdf

What makes a great dissertation?

A document that allows you to claim you …

- … made your mark in your chosen field …
- … fundamentally changed the way something is done …
- … introduced a new concept others continue to "mine" …
- … solved a problem that has plagued people for years …
- … will continue to influence (and be cited by) others.

A truly great dissertation will be accessible to everyone in computer science/engineering, not just to specialists.

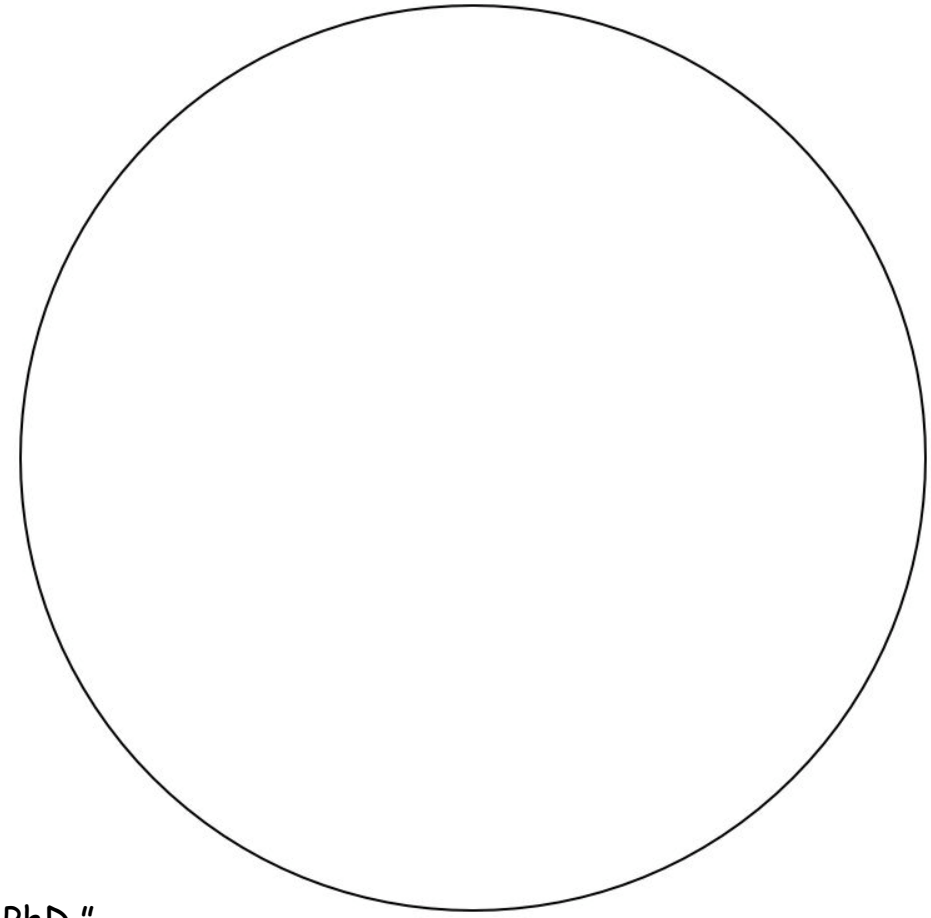
From http://www.cs.cmu.edu/~priya/ICSOC-PhDSymp-2006-dist.pdf

Just to be clear

A dissertation is not a personal voyage of self-discovery.

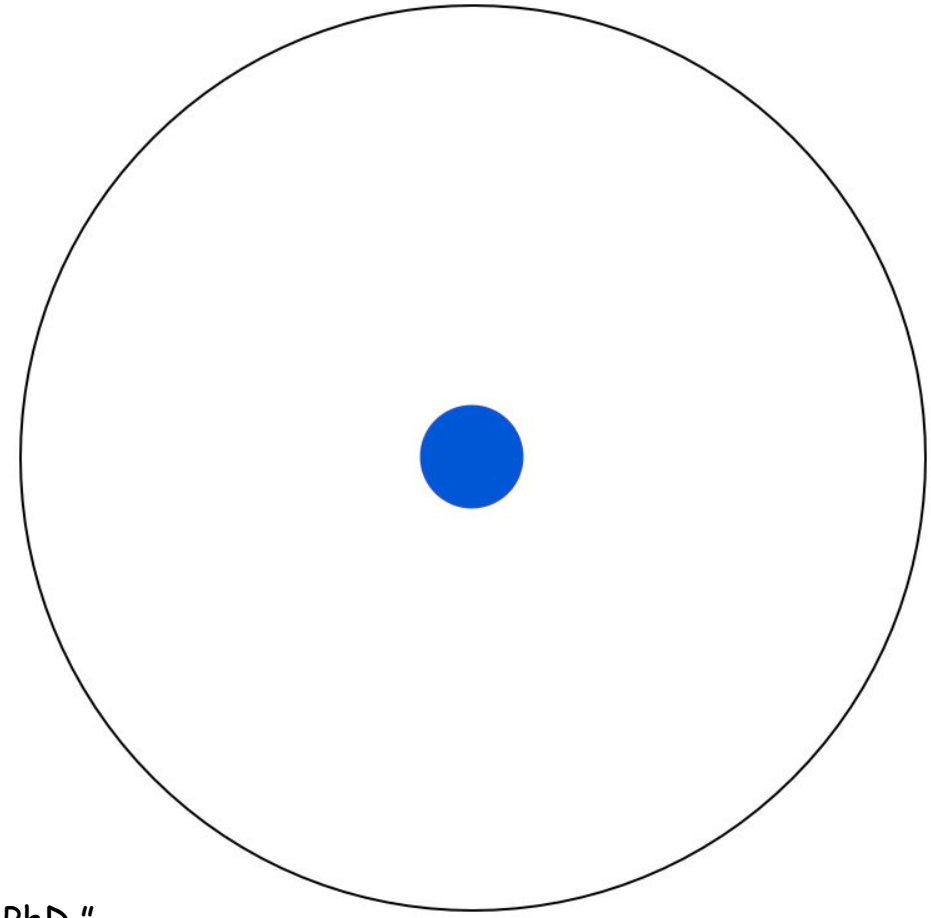
- It is to be assumed you are learning something – that by itself is not sufficient.
- It is to be assumed you are working hard – that by itself is not sufficient.
- The entire field – including the very smartest people who practice in it – must learn something significant in reading your dissertation.
- The world does not need mediocre dissertations that sit on a shelf gathering dust.

This circle contains all of human knowledge



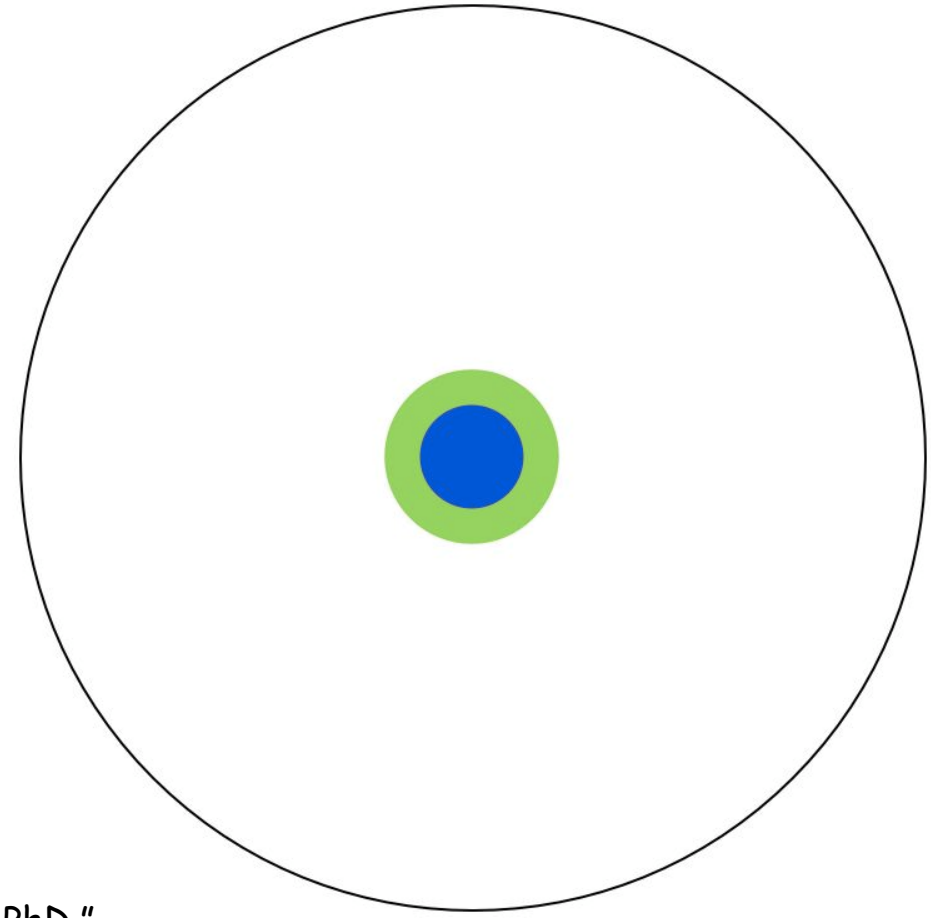
From "Matt Might's Illustrated Guide to a PhD."

When you finish elementary school



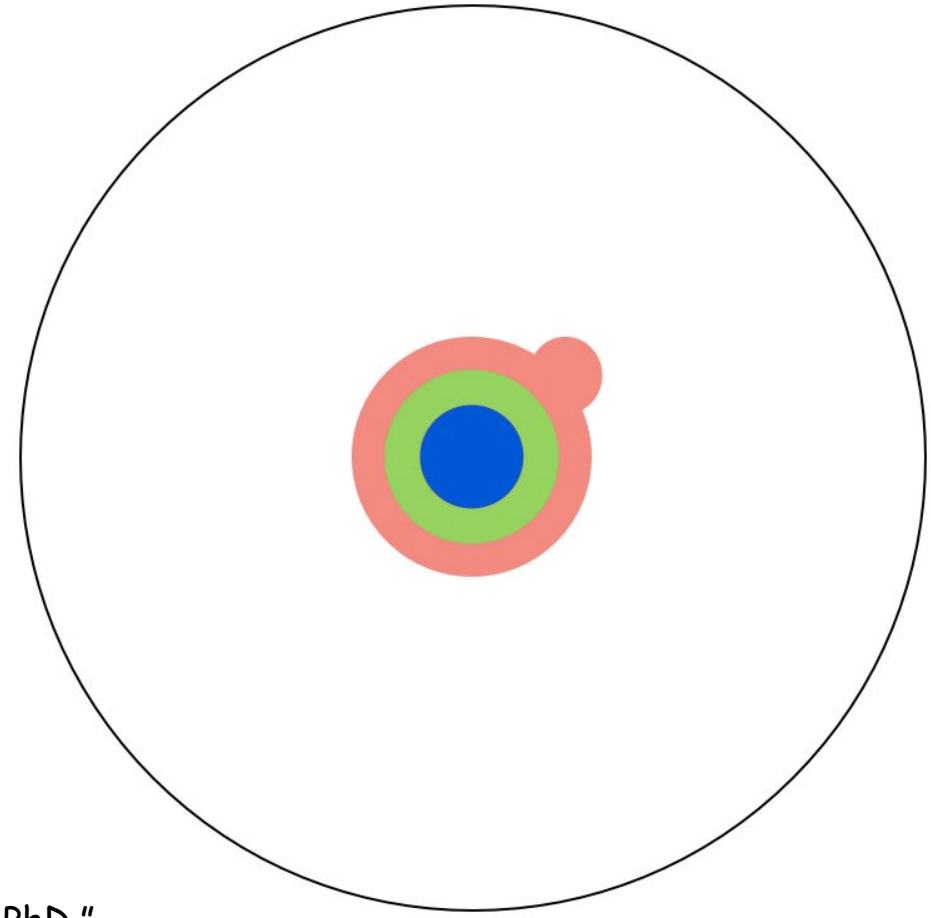
From "Matt Might's Illustrated Guide to a PhD."

When you finish high school



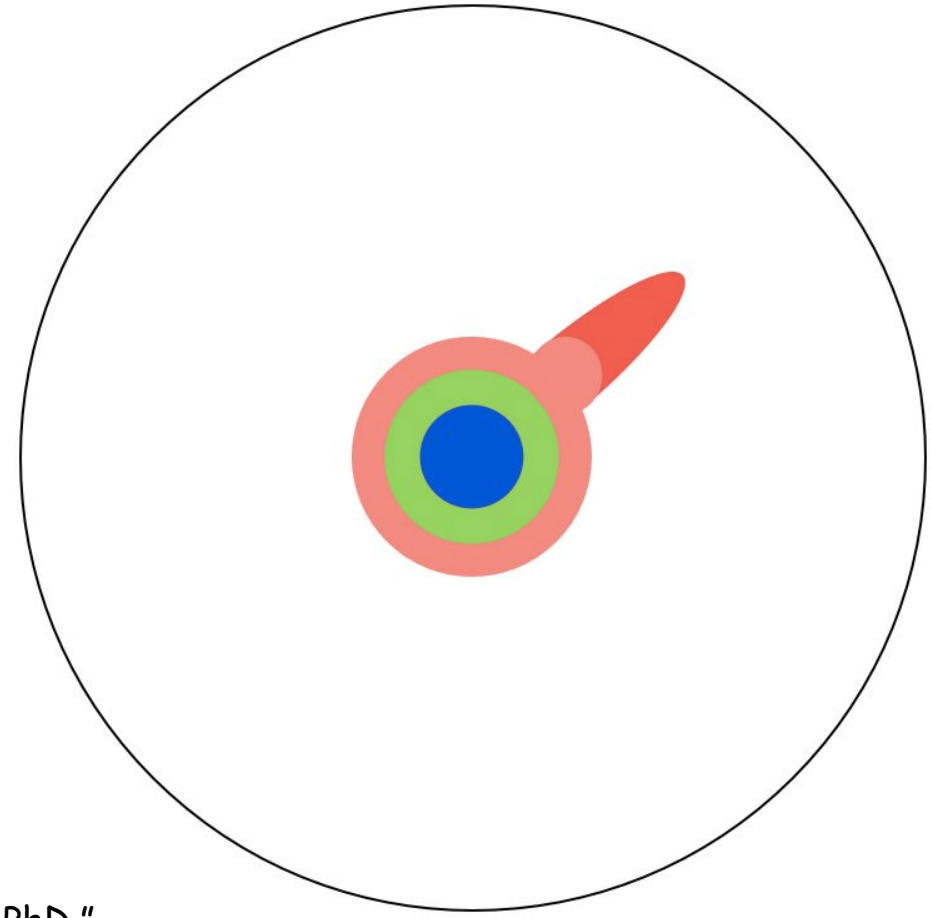
From "Matt Might's Illustrated Guide to a PhD."

When you finish undergrad studies



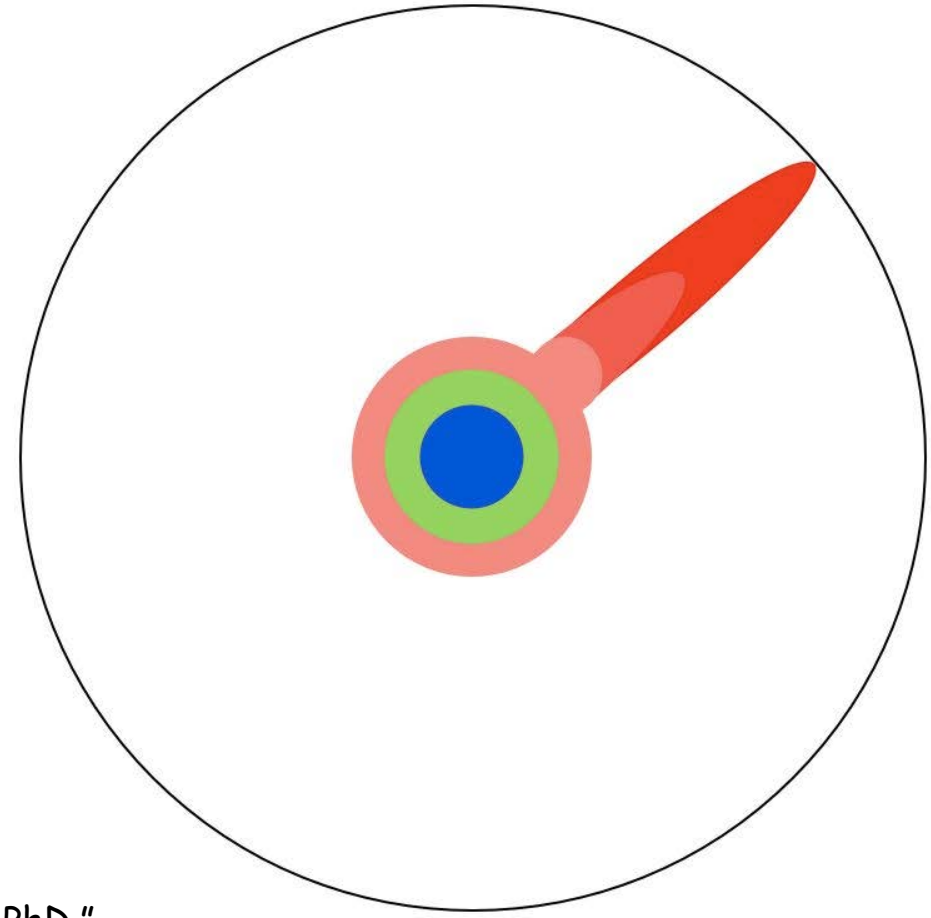
From "Matt Might's Illustrated Guide to a PhD."

When you finish a Master's degree



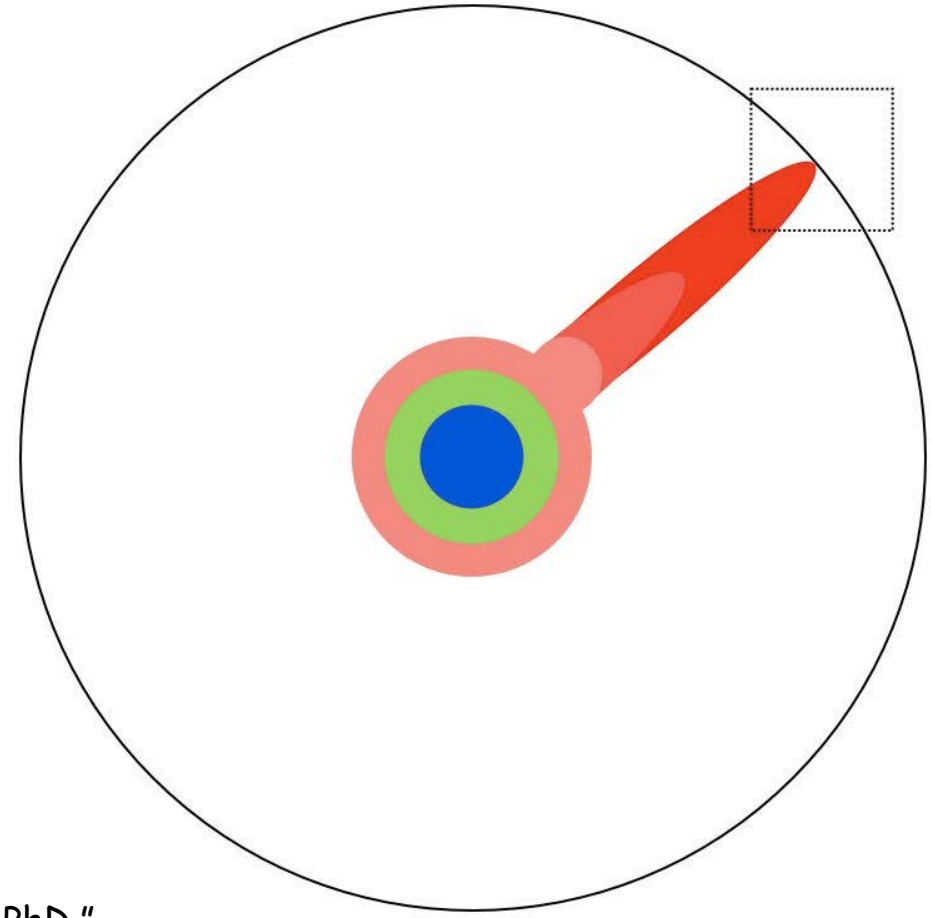
From "Matt Might's Illustrated Guide to a PhD."

Reading research papers takes you to the edge



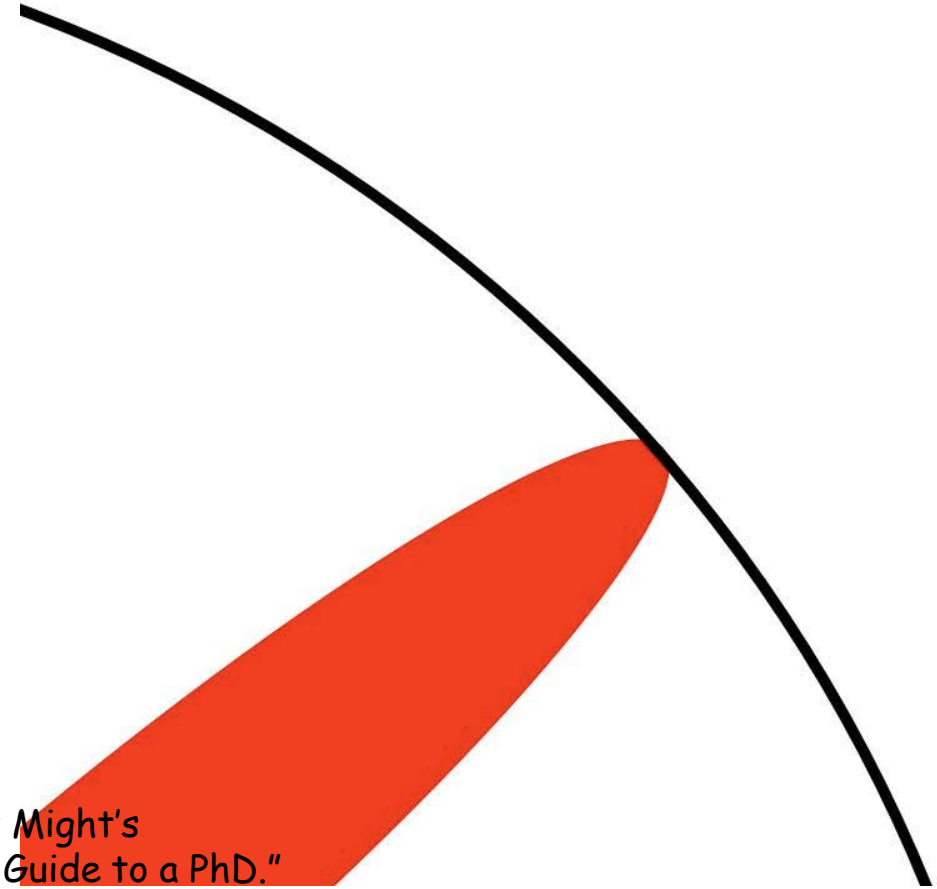
From "Matt Might's Illustrated Guide to a PhD."

At the boundary, you focus



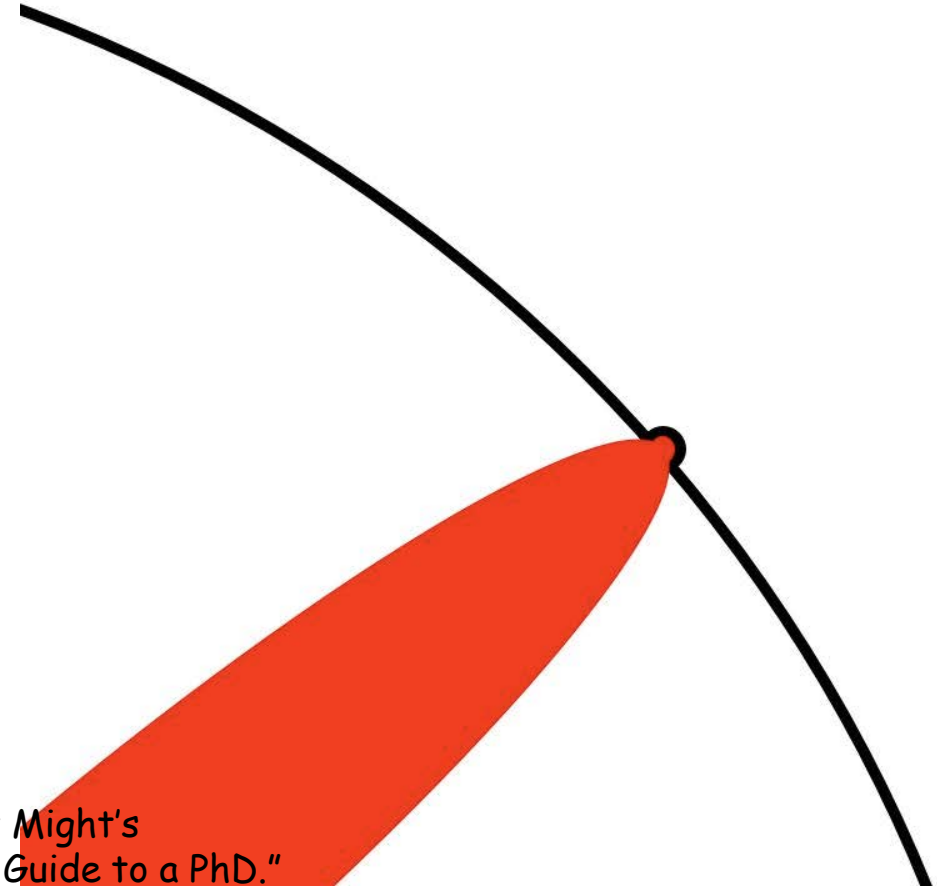
From "Matt Might's Illustrated Guide to a PhD."

You push at the boundary for several years…



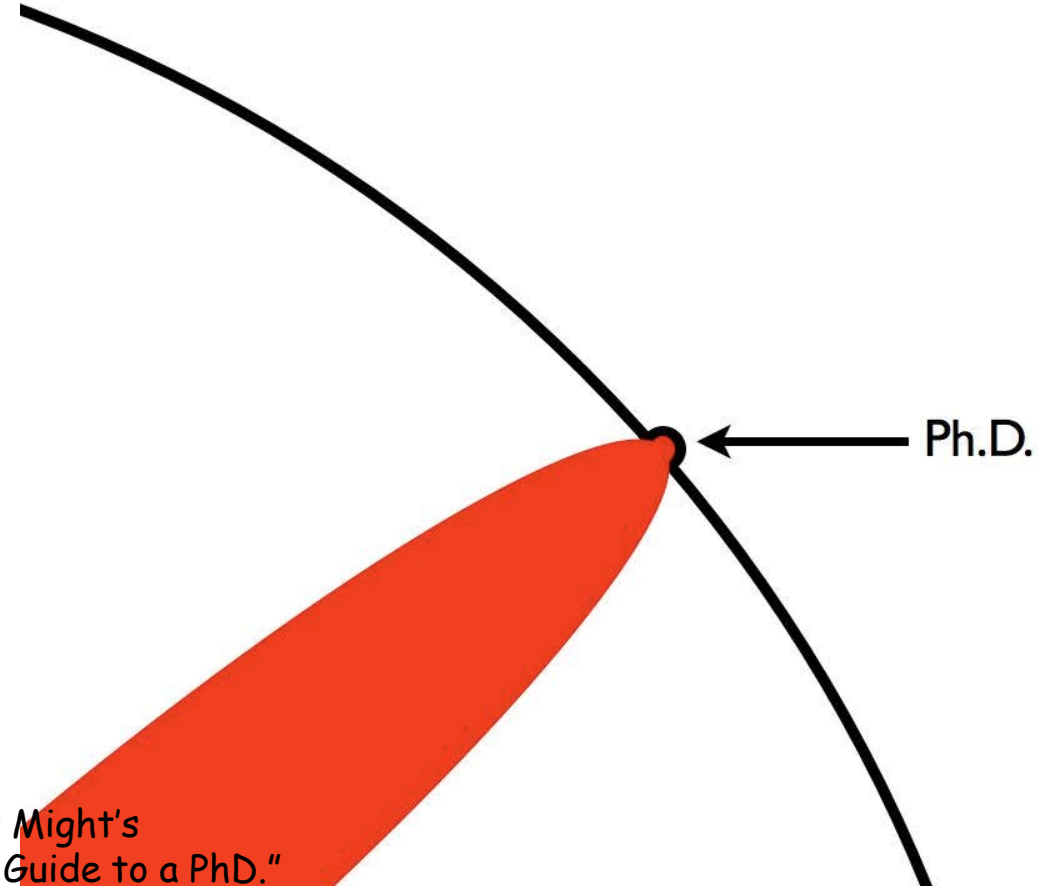
From "Matt Might's Illustrated Guide to a PhD."

One day, the boundary gives way



From "Matt Might's Illustrated Guide to a PhD."

That dent is your PhD



From "Matt Might's Illustrated Guide to a PhD."

The basic idea (1)

- A thesis is a hypothesis or conjecture.
- A PhD dissertation is a lengthy, formal document that argues in defense of a particular thesis. (The term "thesis" is often used to refer to the document, so dictionaries now includes it as one meaning of "thesis.")
- Two important adjectives used to describe a dissertation are "original" and "substantial." The research performed to support a thesis must be both, and the dissertation must prove this is so. A key goal of the dissertation is to highlight its original contributions.

From http://www.erp.wisc.edu/profdev/how_to_write.pdf

The basic idea (2)

- The scientific method means starting with a hypothesis and then collecting evidence to support or disprove it.
- Before you can write a dissertation on a particular thesis, you must collect evidence that supports it.
- Often, the most challenging aspect of writing a dissertation consists of organizing the evidence and associated discussions into a coherent form.

From http://www.erp.wisc.edu/profdev/how_to_write.pdf

The basic idea (3)

- The essence of a dissertation is critical thinking, not experimental data. Analysis and concepts are key.
- A dissertation concentrates on principles: it states the lessons learned, and not merely the facts behind them.
- Every statement must be supported by a reference to the scientific literature or by your own original work.
- A dissertation does not repeat details found in published sources; it uses these results and refers the reader to the original work for further details.

From http://www.erp.wisc.edu/profdev/how_to_write.pdf

The basic idea (4)

- A dissertation must satisfy the stringent rules of formal grammar (e.g., no contractions, no colloquialisms, no slurs, no undefined technical jargon, no hidden jokes, and no slang, even when such terms or phrases are in common use in the spoken language).
- The writing in a dissertation must be crystal clear.
- Shades of meaning matter; the terminology and prose must make fine distinctions.
- Each statement in a dissertation must be correct and defensible in a logical and scientific sense.

From http://www.erp.wisc.edu/profdev/how_to_write.pdf

Definitions and terminology

- Each technical term must be defined by reference to a previously published definition or by a precise, unambiguous definition that appears before it is used.
- Each term should be used in one and only one way throughout the dissertation (no "overloading").
- Best way to avoid a long string of definitions is to say: "the terminology used in this document follows that given in [CITATION]." Then, only define exceptions.
- The introductory chapter can give intuitive definitions of terms provided they are defined more formally later.

From http://www.erp.wisc.edu/profdev/how_to_write.pdf

Results, not circumstances

- "After working eight hours in the lab that night, we realized …" has no place in the dissertation. It does not matter when you realized it or how long you worked to obtain the answer.
- Another example: "Jim and I arrived at the numbers shown in Table 3 by measuring …" You can put an acknowledgement to Jim in the dissertation, but do not include names (even your own) in the main body.

From http://www.erp.wisc.edu/profdev/how_to_write.pdf

Cause and effect

- A dissertation must carefully separate cause-effect relationships from simple statistical correlations.
- For example, even if all computer programs written in Professor X's lab require more memory than the programs written in Professor Y's lab, it may not have anything to do with the professors, or the labs, or the programmers (e.g., it could be that the people in Professor X's lab are working on applications that require more memory than those in Professor Y's lab).

From http://www.erp.wisc.edu/profdev/how_to_write.pdf

Draw only warranted conclusions

- Even if the cause of some phenomenon seems obvious, do not draw conclusions without solid evidence.
- For example, if programs run much slower on computer A than on computer B, you cannot conclude that the processor in A is slower than the processor in B unless you have ruled out all differences in the computers' operating systems, input and output devices, memory size, memory cache, and internal bus bandwidth.
- You must refrain from such judgments until you have conducted carefully controlled experiments.

From http://www.erp.wisc.edu/profdev/how_to_write.pdf

Commerce and science

- In a scientific dissertation, you should never draw conclusions about the economic viability or commercial value of an idea.
- A scientist must remain objective about the merits of his/her work independent of its commercial impact.
- In particular, a scientist never assumes that commercial success is a valid measure of merit.
- Statements like "method Y is used in products by over two hundred vendors" do not belong in a dissertation.

From http://www.erp.wisc.edu/profdev/how_to_write.pdf

Politics and science

- Avoid all political influences when assessing ideas.
- It should not matter, obviously, whether government bodies, political parties, religious groups, or other organizations endorse an idea.
- More subtle and easier to miss, it should not matter whether an idea originated with a scientist who has won the Nobel prize or with a first-year graduate student.
- Always assess an idea independent of its source.

From http://www.erp.wisc.edu/profdev/how_to_write.pdf

Managing your thesis topic

- Your thesis topic is not carved in stone.
- Many students change their topics as they work, paring down certain parts of the project or adding others.
- While you want to keep your advisor and committee informed about major changes, in most cases it is not necessary to be strict in following the research and writing plan you described in your dissertation proposal.
- Research is unpredictable: a smart researcher needs to be flexible and adaptable.

From http://www.unc.edu/depts/wcweb/handouts/dissertation.html

Managing your advisor

- By the time you start working on your thesis in earnest, you should expect to assume some independence.
- By the time you finish your project, you will know more about your subject than your advisor or your committee.
- The student/teacher relationship you have with your advisor will necessarily change as you take this big step toward becoming more of a colleague.
- Believe it or not, this is what your advisor wants to see, just as parents want to see their children succeed.

From http://www.unc.edu/depts/wcweb/handouts/dissertation.html

The big "no-no"

- One subject never open to debate is the level of work necessary for an acceptable dissertation.
- Earning a PhD requires making a significant contribution to the field. Graduate students do not have the perspective to know what qualifies as "significant;" this determination falls within the realm of the advisor.
- By itself, the fact that you believe you have done a lot of work means nothing. When you receive direction from your advisor, you should give it serious consideration and, in nearly every case, act on it.

From http://www.cse.lehigh.edu/~lopresti/Resources/GradSuccess.pdf

Publication

- At the same time you are working on a problem, there is an international research community also working on it.
- Your goal is to gain acceptance into the community by making contributions that are valued by the community.
- How do we determine whether other researchers value your work? By their willingness to include papers you have written in their conferences and journals.
- Substantial portions of your dissertation should be published (or accepted for publication) by the time you defend your thesis.

From http://www.cse.lehigh.edu/~lopresti/Resources/GradSuccess.pdf

Plagiarism

- At this point in your academic life, plagiarism is even more of a serious issue. It could be career-ending.
- You must never attempt to present someone else's ideas or work as though it was your own.
- It is easy to feel time pressure when faced by a paper deadline. Do not give in to temptation and copy text.
- Know that what you do may impact your co-authors, too. I was told a story about a professor who lost his job when one of his students plagiarized in a paper with him.
- Some conferences now use automatic software to check all submissions for plagiarism – you will get caught!

Dissertation anxieties

The dissertation marks the transition from student to scholar and is stressful as a result:

- When you embark on this large, independent project, you may begin to ask yourself questions about your future in academia.
- When you finish your dissertation, you have to change your life pretty dramatically – you may go on the job market, begin work as an independent scholar, develop classes, move out of a community that you have grown to love, and so on.
- You may also feel like your dissertation will begin to define your professional identity. You may feel like your research interests, your theoretical influences, and your skill as a writer may all be evaluated by this first piece of serious scholarship

From http://www.unc.edu/depts/wcweb/handouts/dissertation.html

Maintaining sanity

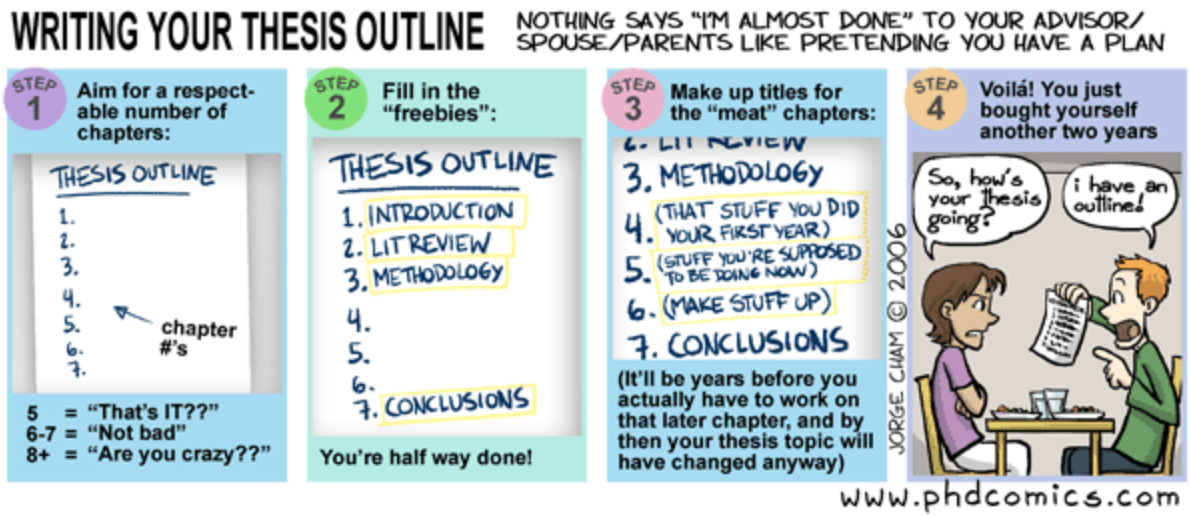
Remember there is no shame in not completing a PhD:

- Many, many people lead happy, fulfilling lives, build lucrative and rewarding careers, make important contributions to knowledge, share interesting ideas with others, and generally get along just fine without the three letters "PhD" after their names.
- Deciding not to continue with a PhD does not mean that you have "quit" or that others who remain in the program are smarter, more driven, or more virtuous than you are.
- It also does not mean that you have wasted the time and money that you invested in the degree up to the ABD stage.
- It simply means you decided this career choice was not for you.

From http://www.unc.edu/depts/wcweb/handouts/dissertation.html

Conclusion

Best Wishes for Success in Your PhD Studies!



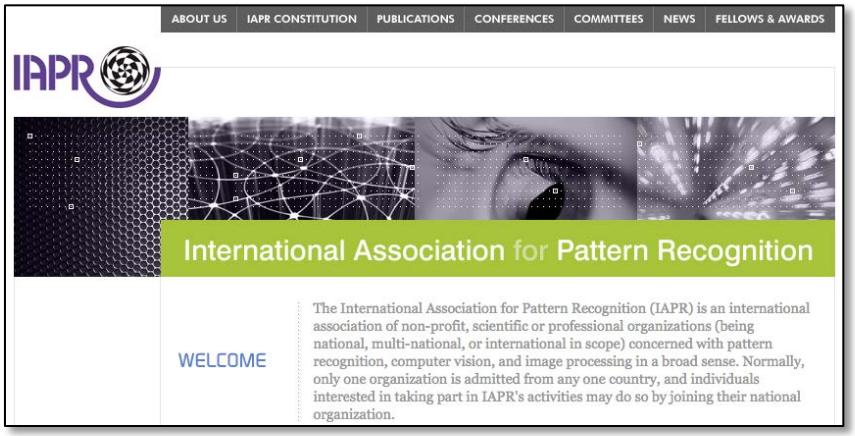
When you need a dose of humor: http://www.phdcomics.com/comics.php

A Few More Good Things for You to Know

IAPR (www.iapr.org)

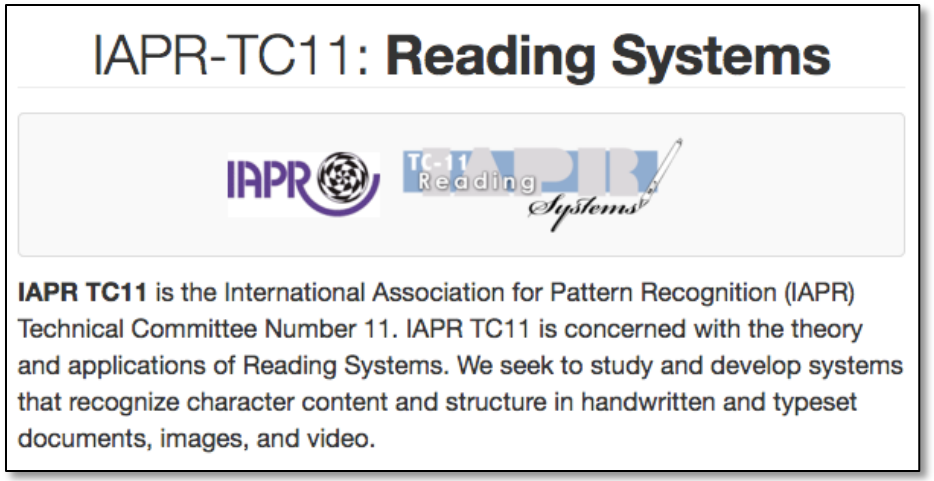
IAPR – the International Association for Pattern Recognition – is the organization responsible for supporting a wide range of activities by the research community.

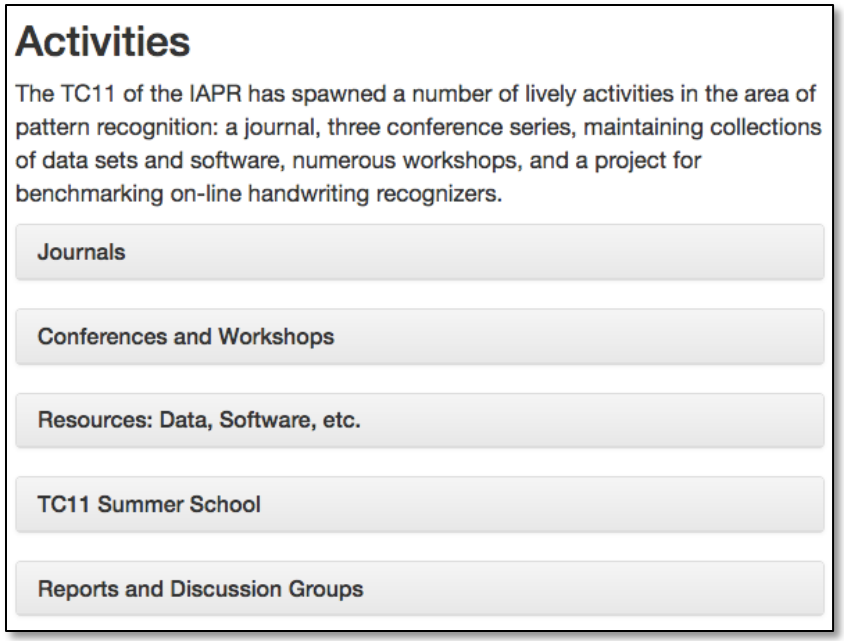
- IAPR organizes conferences, administers awards.
- IAPR technical committees do much of the actual work …
- TC-10 (Graphics Recognition),
- TC-11 (Reading Systems).



IAPR TC-11

TC-11 ("Reading Systems") represents the international research community in topics relating to character recognition and document analysis.

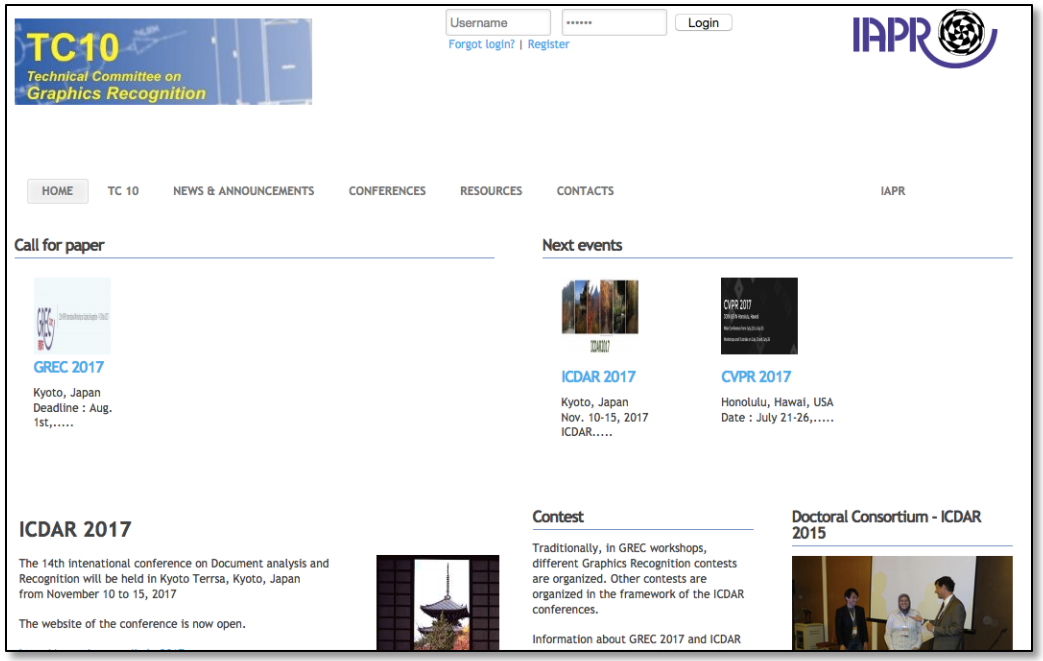




www.iapr-tc11.org

IAPR TC-10

Graphics Recognition is an exciting field of pattern recognition. IAPR's Technical Committee 10 on Graphics Recognition promotes interaction among researchers working in document image analysis in general, and graphics recognition in particular.



iapr-tc10.univ-lr.fr/

IAPR TC-10 and TC-11 Resources

As a beginning researcher in document analysis, you should:

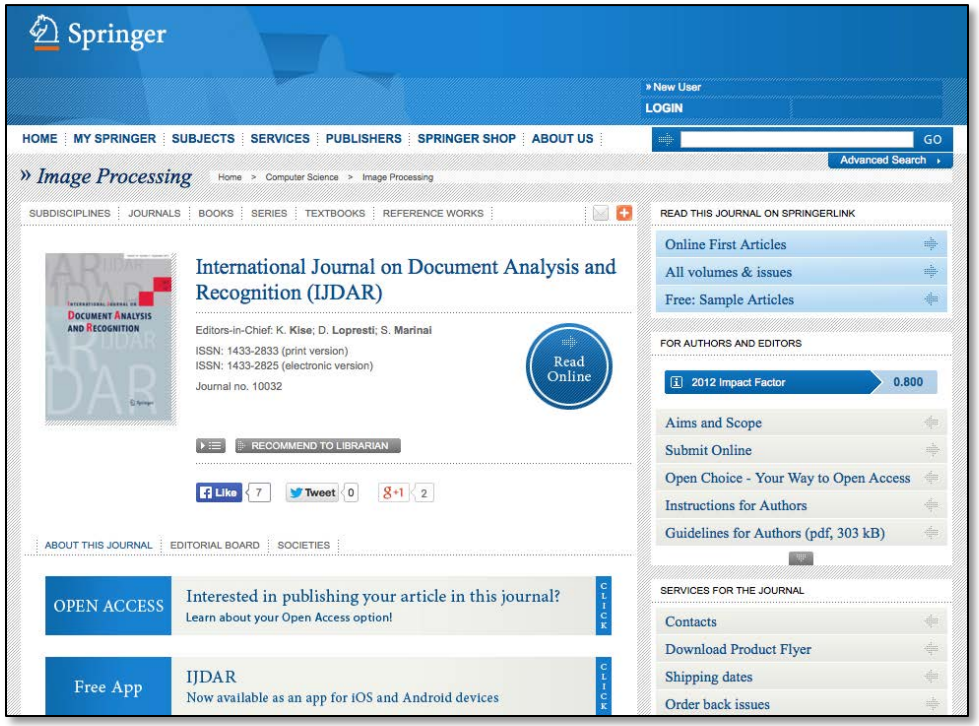
- Study the resources on the TC-10 and TC-11 websites: in particular, the conferences and the datasets.
- Do not invent a new dataset if an appropriate one already exists. Even if you believe the problem you are working on is "unique," try hard to find some way to compare your methods to existing ones using a known dataset.
- If you create a new dataset, contribute it to the international research community.
- Sign up for the TC-10 and TC-11 mailing lists!

Upcoming IAPR Conferences

- 13th International Workshop on Document Analysis Systems (DAS 2018), April 2018, Vienna, Austria.
- 16th International Conference on Frontiers in Handwriting Recognition (ICFHR 2018), August 2018, Niagara Falls, NY.
- 24th International Conference on Pattern Recognition (ICPR 2018), August 2018, Beijing, China.
- 15th International Conference on Document Analysis and Recognition (ICDAR 2019), Australia.
- 17th International Conference on Frontiers in Handwriting Recognition (ICFHR 2020), Germany.
- 16th International Conference on Document Analysis and Recognition (ICDAR 2021), ???

IJDAR

IJDAR – the International Journal on Document Analysis and Recognition (Springer) – is devoted to our field.



- Other fine journals and conferences exist, too.
- Look where your advisor and respected senior researchers publish.
- Aim for high quality.
- Know the literature before you start work.

http://www.springer.com/computer/image+processing/journal/10032

Thank you, and good luck!

ありがとう、そして幸運！