

Writing an abstract with keywords

Hello! This is Gerald Maguire. I would like to welcome to this very brief module on writing an abstract with keywords.

Slide 2: What is the purpose of an abstract?

The first thing to understand is what's the purpose of an abstract. The key idea is that you want to briefly summarize your document. You want to help the reader decide: Should they bother reading your paper or not? The third thing is you want to help potential readers to find your document. Because your document is going to be indexed by search engines based typically on the titles and abstracts (in older databases) and nowadays that many full-text database exist. But primarily, the abstract is going to be a major reason why people find your paper and decide to read it. And of course, one of the goals is that you want to make your research visible to others; so, therefore, it is in your own interest to write a good abstract. Now the abstract is somewhat usual in that it stands on its own. Many times abstracts appear without the full publication, so the reader should be able to read the abstract itself. Therefore, you don't cite your sources, you don't have cross-references into the body of it, and any acronym has to be spelled out in the abstract - because otherwise, the person isn't going to be able to see it because it's (yes) physically partitioned away from the paper. And, of course, it's important to remember that many readers will only ever read your abstract, and then they will decide, "no, this is not for me", and then they'll go on and look at something else. So you want to be honest, you want to make it clear why a reader should read it, but you don't want to mislead them by having them say, "Oh! Yes, this is really very interesting," and then they read it the actual paper and find "no, that doesn't actually contain any of the things I expected it to contain" Then they're going to be disappointed, and when they next see an article or paper by you they'll say, "Nah!, I not going to be even bother".

Slide 3: What are the components of an abstract?

So, what are the components of the abstract? One of the first things is, of course, that it has a motivation. And that typically sets the context. Now, this is often optional, because in many cases, the context may be apparent because it's a paper in a special issue of papers all about the same topic or at a conference in a track where all the papers are about a particular area. But every abstract has a problem statement. What is the problem you're trying to solve? It also says what are the methods and approaches used to solve the problem, typically [including] why you are able to solve the problem whereas others haven't been. You present your results and, of course, your conclusions. And Philip Koopman has a very nice essay called "How to write an abstract" he wrote in 1997 at Carnegie Mellon University, where he was on the faculty, that I would encourage you to read.

Slide 4: What is in a Master's thesis abstract?

So let's take a look at a very particular kind of abstract, a master's thesis abstract since most of you are going to be writing a master's thesis within the next two years. What is in it? Well. The first thing, as we said, is the motivation, What's the topic area? (to set the context for the reader) and a short problem statement. Both of those are to highlight - What is the problem that you're trying to solve. The next part, that is very essential in a master's thesis abstract, is: What makes this worth a master's thesis? So, why is it both significant and a suitable degree of difficulty for a master's thesis? And why hasn't anyone else solved it yet? That is about motivation. The next part is the methods. How did you solve the problem? What was your path or insight that led you to solve the problem? And of course the results, your conclusions. What are the consequences? What's the impact? What is it that you want others to do? What can be done now after your thesis that couldn't be done when you started? And typically, on the same page with the abstract, it's going to be a short list of keywords.

Slide 5: How long is a thesis abstract?

So, how long is a thesis abstract? Students ask me, so I went and looked at the database of past theses, and I found out that for myself, for theses written in English with an abstract in English - had a typical median length of 289 words. The Swedish abstract was about 207 [words]. The minimum length was 100 and 98 [respectively]; and the maximum length was about 606 and 509 words [respectively]. Here are similar statistics for all of the other examiners that were in ICT in this period from 2010 to June 2014. About 90% of of all the English abstracts for theses for my students are between 200 and 450 words, and that matches pretty well with ProQuest Dissertations & Theses (PQDT), which is a database of theses and dissertations, which limits the abstracts to 350 words And in Canada the National Archive limits Master's theses to 150 words doctoral dissertations to 350 [words]. Chalmers University of Technology, here in Sweden, suggests a master's thesis abstract should be between and 350 words - so the length of these abstracts of students that I've had are really well within the range you would expect.

Slide 6: Histogram of number of words in ICT thesis abstracts

Now is it just about the number of words? No, of course, it is about the content. But if you look at the distribution of their lengths, we find they are very few that were long, very few that were very short, and most of them are in this range. There are several hundred theses there, and it's probably statistically significant, [hence] that's a reasonable range.

Slide 7: At KTH, every thesis must have an abstract in both Swedish and English

Now, as you are a student at KTH, it important to know that at KTH, every thesis **must** have an abstract in both Swedish and English. And this is spelled out in the "Language Policy for KTH 2020". This means that, of course, yes - if you write your thesis in English, you still have to write a Swedish abstract, or if you write your thesis in Swedish, you still have to

write an English abstract. Now, one fortunate thing is that additionally, all theses are reported in the KTH's digital scientific archive (DiVA) - since a decision in January 2011. So, of course, it's possible to go and see all of those different abstracts, find them, match them; I even have built a tool to take the keywords and the words of abstracts and compute matching words and come up with vocabulary lists.

Slide 8: Keywords

So, I mention keywords. One of the reasons to have a set of keywords with your article or your publication is to help readers find it. That means you need to think of what are the most likely keywords that a potential reader is going to use. And how do I do that? Do I just think of words? No, I can go to entities like the ACM Computing Classification System or the IEEE Taxonomy - where they, in fact, give me - well-defined keywords for particular areas. I can choose from that. The mechanics of the list of keywords is the first letter of the keyword should be set with a capital letter, and a proper name should be capitalized as usual. We spell out acronyms and abbreviations. Because, of course, it isn't necessarily the case that the hash is one way - there are many areas that use the same acronym and abbreviation, and we want our readers to know which particular meaning it is. We get rid of stop words because they contain little or no information. And we separate the keywords by commas. And I suggest that since you're going to have both English language and Swedish [language] keywords, that you choose one order and you make the other one match in that particular order - that way it's mechanically possible to go and find the matching keywords.

Slide 9: New Keyword(s)?

But sometimes you're going to find out there isn't an existing keyword in one of these taxonomies or earlier definitions, that means that (yes) you need to introduce a new keyword - but you should think very carefully about introducing it. So, for example, L. Wilson Pearson, who is the editor in chief of IEEE Transactions on Antennas and Propagation, in the two and a half years that he was editor of the journal - only added a 150 new keywords. Why is it a bad idea to add a new keyword? Because it's unlikely that someone else is going to think about this keyword. Of course, if you introduce the landmark paper in an area, then you might (yes) introduce a new keyword, and then it becomes something that everyone is going to search for. And (yes) they're going to find your document and the others who followed.

Slide 10: How many keywords?

So how many keywords? Well, for the theses in DiVA in that period of 2010-June 2014, the median number of keywords was 5, with a minimum of 1, a maximum of 43 in English and a 146 in Swedish. The minimum and maximum standard deviation ranged from 3 to 5. That means a typical thesis is going to have 5 to 6 keywords. And that is not so surprising, because most publications enforce a limit of 4 to 10 keywords.

Slide 11: Keywords in ICT theses (2010-June 2014)

So, if you look at the number of keywords in ICT theses, here's the distribution for them. With these little diamonds, we can see the distribution in English and the frequency, and we see that it has a high of 4. We see 3, 4, 5, and 6 are pretty high, all the rest are pretty low. The frequency for Swedish keywords, the numbers of them because there's fewer of them, are again in about that 3 to 6 range. And (yes, there are some that have very large numbers). But 3 to 6 is pretty common.

Slide 12: Collected keywords

Now, one of the exercises that I did was I actually looked at the collected keywords used in different parts of KTH. So we can see in the left-most column here, we can see the theses from ICT, and here are the major keywords that were used. And I use this vertical bar "s" to indicate "system or systems" And if we look at this top set of keywords these are the most frequently appearing keywords in ICT and compare them to EES or CSC, relatively similar areas, we have "system|s" at the top for all of three of these, but then there's a big difference that goes "networks|", "model|s", "method|s", and for ITM "energy" is now the [second] third biggest one. So, we can see that there is quite a big difference between the different parts of KTH. in the keywords that are used by theses in those areas.

Slide 13: Vocabulary list

As I mentioned, I mined the information from the keywords and from some of the courses I teach, and I was able to produce a vocabulary list and with the help of Markus Hidell and various dictionaries to create a pair keyword list. And we see that for ICT they're about 3853 pairs between English and Swedish and that is the basic vocabulary that you need to know. So, if you're saying, "Oh! But I have to write the abstract in Swedish" - well, one can find out fairly easily much of the technical vocabulary that you're going to need for those matching words - from such a vocabulary list.

Slide 14: References

Their number of references you can read them to learn more about writing an abstract and selecting your keywords. I wish you good luck in doing both of those.