Proposal to limit MSc project report length

Sharon Goldwater, 11 Nov 2018

Summary

The problem:

- We want to improve the quality of dissertations and the quality of feedback we can give on dissertations while coping with large numbers of dissertations (perhaps as many as 6-7 per supervisor this year, with a marking load of perhaps 15).
- The MSc project pages say that project report length "should be 40-50 pages in total, and no shorter than 20 pages". In practice, the average length is about 48 content pages[1], with many reports significantly longer. This suggests that students believe that 40-50 pages is 'expected', and to 'excel', the report must be even longer. The guidance thus encourages a focus on quantity (of work and of writing) over quality, and also makes it difficult to provide high-quality feedback and timely marking.

Solution idea and justification:

We should follow the lead of many other universities and Schools within UoE, and impose length limits on MSc dissertations. This will
- encourage students to focus more on relevance and quality than on quantity;
- encourage students to focus more on their own contributions and less on background material (which they have often already presented in their IPP anyway);
- make it easier for supervisors to provide quality feedback on drafts; and
- enable more thorough and consistent marking of final submissions.

Specific policy proposal:

- Impose a hard limit of *40 content pages*[1] on MSc project reports, assuming the current Infthesis MSc format with 1.5 spacing. (The use of the InfThesis style is also required.)
- No exceptions are permitted. Appendices are allowed, but it cannot be assumed that the marker will read them.
- Project reports will not be accepted if they do not conform to the length limit. This can be achieved through a modified submission process (see below).
- To avoid confusion, the policy applies to all 60-credit MSc dissertations administered by the School, excluding those from EPCC: the MSc Dissertation (Informatics), DSTI Dissertation (Distance Learning), and Masters Dissertation (Design Informatics).[2]
- Since the credit value of CDT dissertations varies, they are not included in the proposed policy, but could be if requested by CDT directors, probably with a different length limit.

Implementation details

No exceptions to the page limit or InfThesis style are permitted, otherwise some students have an unfair advantage over others. The following implementation is proposed, with MScPM taking responsibility. Some items will require liaison with ITO and/or computing support.

---

[1] We define the content pages to be the introduction through conclusion, not including front matter, bibliography, and appendices.

[2] In principle, Design Informatics could be treated differently, but as noted this is likely to cause confusion.

- Strictly enforce the InfThesis style. This requires clarifying instructions, including how to use the DICE-installed template, how to install the template on non-DICE systems (e.g., Overleaf), and clear specs/sample document for non-LaTeX users. Deadline: 1 Feb.
- Modify the submission form so that at submission time, students are required to self-certify that their report conforms to the correct style and length. (This should be a minor modification to the form, which already requires students to confirm that all work is their own.) They should be warned (in advance and on the form) that a later check will take place and the submission may be rejected if it is found to violate the requirements.
- Immediately following the submission deadline, MScPM uses a script to process the submissions[3], yielding: (1) a small list of submissions flagged as possibly too long, and (2) a single pdf document consisting of one page from each submission, watermarked with the student number. These outputs are sent to ITO, who then visually scan through the pages from (2) to identify any style violations, and double check any dissertations flagged by (1).
- The authors of any reports that are found to violate the requirements will be notified by ITO and told that they must fix the problem. Standard late penalties will apply to resubmissions.
- Provide clear guidance to students and markers regarding this policy, frequently and starting in the project proposal phase. Emphasize that the dissertation quality depends on a good balance of background information and new material, and deciding what to include or not.
- Provide at least one example of a high-quality dissertation that falls within the length limits. (I can point to one 35-page example that received an 80 on a PPLS degree, but would easily qualify for an Informatics degree: the topic is unsupervised learning for speech processing.)

Justification

Why not a word limit?

Some colleagues have suggested that a word limit would be better, since some topics may require more visuals than others. However, word limits are not typical of our field and are hard to verify. They also require difficult decisions about what counts as a word (numbers in tables? equations? etc) and trusting students to use word-counting software that is consistent with these decisions.

Students should be able to choose key visuals with discretion, and supporting visuals could be put in the appendix. To the extent that a student chooses to omit necessary visuals in order to fit more words, this would be penalized via the 'quality of the dissertation' marking criterion.

Why 40 pages?

Discussion with a handful of staff suggests this is not an unreasonable choice. I based it on my experience marking many dissertations from Informatics and from PPLS (normally on NLP topics), where a 10,000-word limit is imposed (equivalent to about 32-35 content pages in Infthesis format). I find that the PPLS dissertations are typically much more focused and concise, but a bit tight on space for implementation details. On the other hand, most Informatics dissertations of 45-50 pages (or longer) contain unnecessary background material and/or a parade of experiments which could benefit from being trimmed down to focus on the most interesting/relevant ones. A 40 page limit would force some students to edit, while still permitting enough space to describe necessary implementation details. It would likely also reduce the amount of material replicated from the IPP.

---

[3] Similar scripts are used by conference publication chairs to check formatting and length. Iain Murray has tools that do essentially what is described here. The length check uses heuristics and isn't perfect, but is good enough that only a few documents would need to be checked by hand. It is estimated that no more than a few hours of work would be needed to convert these scripts for use with MSc project reports.

Competitor analysis

In a quick search of other top UK MSc programmes, several have length limits that are not out of line with those proposed here. In particular,

- The UCL Software Engineering MSc project report has a limit of 12 pages in ACM SIGCOM conference format.
- Several Cambridge MPhil degrees (Machine Learning, Speech and Language Processing, and Advanced Computer Science) all have 15,000-word limits, which would typically be only slightly longer than the proposal here.

Another data point, though not from a computing degree: a colleague in biology at Imperial reports that they have a 50-page limit, including front and back matter, on their 90-credit MScR. Our MSc thesis is only 60 credits and the 40-page limit proposed here excludes front and back matter.)

Why not just use a penalty?

Some staff suggested that over-long reports should be accepted, but a penalty imposed (for example, limiting the maximum mark or subtracting a percentage of the mark). This idea was considered and rejected for the following reasons:

- it's not clear what the correct penalty/penalty structure should be;
- it leaves checking the length and perhaps even deciding what to do about it to individual markers, which will lead to inconsistency (and more work for markers);
- social science evidence suggests that allowing such submissions with a penalty may actually lead to more students choosing to violate the policy: the opposite of the desired effect.

Consultation

Consultation to date:

An initial version of this proposal was tabled at Teaching Committee on 12 Sep, and included both UG and MSc projects in the proposal. Most staff in attendance favoured some form of length limits, but some concerns were raised about the lack of writing support that UG students receive, and the fact that length limits on UG dissertations might prevent some of our students from winning national prizes. The latter concern is irrelevant for MSc, and the former is mitigated by the fact that MSc students receive feedback on writing during IRR, IPP, and many of the largest MSc courses.

The previous proposal did not include a concrete mechanism for implementing the requirement. This has now been added based on discussion with a few members of staff.

Suggested path forward:

I suggest we use the TC meeting to decide whether the implementation (below) is acceptable in principle if the policy is adopted (regardless of whether members who attend agree with the policy itself). If so, then the wider staff can be consulted by email on whether to approve the policy. We may wish to allow only staff who are actually required to supervise MSc projects to weigh in on the decision (e.g., those who have had this as a duty in the previous two years or this year).