Notes Nov 2019

Abstract: it is likely that the abstract would be read by people with no idea about MDE. Put in a short first sentence, and revise lines 16-20: something like:

> In large-scale computer systems and software development, model-driven engineering is an approach that focuses on the development and management of models -- design diagrams, text models, code. Tool-supported model management is well researched in industry and academia. A key bottleneck in model management is the time taken to load models into model management tools. In this thesis, we present an alternative to traditional state-based model persistence, referred to as change-based persistence. Persisting…

Otherwise fine, and preamble pages are fine too.

Chapter 1

Line 259: replace "this work" with something like "to the research presented in this thesis".

L 260: present tense is best for a thesis: the research questions that are addressed by the research.

§1.1: as in the abstract, start with a minimal introduction to MDE (see suggestion above).

L270: "computationally consuming" -> computationally expensive / inefficient
L270-1: perhaps a better last sentence would be something like:
> The research presented in this thesis is motivated by the need to find a more efficient approach to model differencing and conflict detection.

L272-3: remove "also" and "instead" (you said it's an alternative; you don't need other words that say the same).

L276-7: perhaps better as:
> Change based approaches facilitate detection of differences between versions, and makes better semantic identification of differences -- …
L278: Break the sentence at the dash, and expand. eg.
> etc.). Better and more granular identification of differences allows better support for resolution of conflicts e.g. where versions of a model have been modified in different ways.

L293: again, present tense better, but could improve wording, e.g.
> Every time a model is modified, the file of changes increases in size.

L295: present tense

L298: start a new paragraph: suggest rephrasing eg.

One approach to controlling the size-increase of the change file is to identifying changes that do not affect the eventual state of the model. Loading time can be improved by ignoring -- i.e. not replaying -- changes that cancel out.

§1.2

L315: start a new paragraph after the hypothesis statement.

L317-322: not well stated. I think you mean something like:
Model differencing is used to identify how two or more versions of a model differ -- for example, determining what has been changed from the original model, or comparing new versions of a model created by different teams working independently. A typical goal of model differencing is to ascertain whether independent updates can be merged, or whether there are conflicts in the model versions: elements or features that differ in ways that are incompatible. Conflict resolution requires identification of conflicts, and user intervention to decide which the appropriate state of the conflicted elements of features in the eventual merged model.

Somewhere here, you need to mention "model" definition as well… e.g.:
The hypothesis models are, here, typically graphical models that capture the structure of the data or behaviour of systems, though, in principle, the research can use any models that needs to be efficiently stored, loaded and compared.

L322: consider new paragraph before "This work" and sort out wording, e.g.
The hypothesis refers to large models. In this research, a large model is taken to have more than one million elements, a definition that is consistent with [10,11].

L336-8 Phrasing RQs is hard: is this better?:
1. How can models be effectively persisted in a change-based format, and how does change-based persistence perform, compared to… … (RQ1)
   ...
2. In a changed-based format, how can differences between models be identified, and how does change-based model differencing perform, in terms of … (RQ2)
   …
3. Following change-based model differencing, how can conflicts be detected between versions of a model, and how does change-based conflict detection … (RQ3)

On RQ1: will you talk about how models are represented in tools (or at least reference model expression languages) later in the thesis? May be worth a note here e.g.
(the languages used by tools to represent models are summarised in Section x.y.)

On all the RQs, presumably you also investigate whether the change-based persistence approaches are at least as accurate as the state-based ones in terms of reconstructing a model version, differencing models and conflict detection?

§1.3

Objective 1 (l363): probably better to say "Research and design an implementation ..."

I wonder if it is worth noting some of the other things that you need, such as very large models and a corpus of large models… so an early objectives might include to identify large model examples to use for analysis of the cbp performance; to generate large models for use in research on representation and use of cpb...

§1.4

Prototypes of what? Prototype implementations of change based persistence (storage, model loading, and mode-management operations)?

Check with Dimitris: I don't think that publications and a thesis are really research outputs as such… but the analysis experiments and results are.

§1.5

Point 1: it would be nice to know why… at least a forward reference to where this is discussed.

Better to talk about merging etc. as being "beyond the scope of the research presented in this thesis" and "constitute natural extensions and further work".

§1.7

A better lead in would be:
> The research in various parts of the thesis has been published in the following papers:

For each paper note where the (main) material is in the thesis: e.g.
> (Chapter X and sections Y.Z-Y.A)

Overall, the chapter is well structured and clear.

CHAPTER 2

L492 ff: weak intro paragraph. Better to note what the review does… e.g. first it presents work on model persistence, highlighting common features and aspects relevant to model persistence. Then it introduces EMF store, because...

§2.1
As you can guess from the above, I expected this to be a lit review of persistence… e.g. OODB etc… but the section strats with a description of model persistence which is your research not review.

I think ll497-506 are largely redundant… but it depends how you structure the chapter… for instance, if you say,

>Chapter 2 starts by considering the ways in which computer science research and practice has presented change-based persistence… ***

… then you can just promote §2.1.1 to §2.1

§2.2

Again, needs something in the chapter intro that leads the reader to expect this e.g.

>*** … , and specifically introduces EMFStoreL a rare example of change-based persistence in MDE.

Ll543-7: clumsy. Possibly:

>EMFStore [9] appears to be the only current implementation of change-based persistence for MDE models, providing a repository and version control management for EMF-based models. Rather than using a general repository such as SVN [ref] or Git [ref], EMF provides specialised model storage with its own change-based…

You need some citations on the information in this section… where does the information you present come from? You can probably reduce the detail… but you also need some value-add (rather than just describing what it does): what is good, what is limited, what can you use, what is a compromise, why can't you just extend EMFStore. [a little bit of this is in ll580ff]
Is it open source?!

Tables 2.1 and 2.2 are a long way from their text reference (l658): can you force the formatting?

Again, §2.1.2 does not seem to follow logically and needs some introduction… it feels by now that you review will almost entirely focus on model persistence…?

I think you need to map (perhaps literally with lines and boxes) the lit review and get a clear "story" structure… then describe the structure at the start of the chapter… tell the reader how you will lead them through related literature etc.

The content is generally OK… not sure if it's been proof read… there are places where it could be phrased better. I didn't spot anything that is obviously missing (examiners nearly always spot something you could add to such a review, though!).

It would be useful to have a short summary/conclusion section to the chapter, so you don't leave the reader thinking "so what?".

CHAPTER 3

Needs a better title… there is a very similar title in chapter 2. Here, it is Designing Change-based Persistence for Models?

Intro could note what in particular you use from the reviewed material (or that could be in the last section of chapter 2.

A box and line diagram is often useful to show the steps that the chapter moves through in the design.

§3.1 seems to be another text motivating the need for cbp… needs reworking with the repetition removing and the parts that are not repetitive putting with other motivating text explaining why cbp is useful and relevant.

Ll865-869 are ok, but explain why the running example is necessary and why it's appropriate (what features you needed to develop the representation etc.).

§3.2
Content is ok but structure is odd: I want to know first what the running example consists of… so a set of models that are versions created from a common original model. I want to know that he diagrams describe the class structure of a role-play game, developed by independent teams from the common model. Then, I want to know about the notations… these are not consistent class models as there are classes that do not have associations… so you need to explain that only the parts of the diagrams that are pertinent to the persistence discussion are shown e.g. in figure 3.1. A short text summary is useful at the start or after Fig 3.1. Is introduced e.g.:

> At the start of development, Jane produced classes for Character and 4 specific types of character, but did not determine the relationships among these. Working separately on this original model, Bob and Alice made different decisions about subclassing, renamed some character types, and, in Bob's case, deleted one character type. The models are still incomplete, and it is desirable to merge the designs at this point. However, the revisions have introduced inconsistencies that we need to identify, and resolve, before creating a new definitive model from the two new versions.

Omit lines 871-2: this belongs in the chapter intro and/or end of the previous subsection.

Could the detailed /complete listings go in an appendix, with just the first few lines of each included, perhaps side-by-side here? It's quite hard to follow the text with lots of listings interspersed.

§3.4

You describe your prototype appropriately, but I feel that it needs some more abstract intuition and rationale… e.g. Fig 3.2: how do you know that these are all the relevant reference events, attribute events etc? Could you write a "requirements summary": e.g.

> From the text-based change summaries in the above listings, we see that we need to be able to identify different features of the models and the different ways that instances of these features can be changed. Because our approach focuses on changes, we identify classes of event, which inherit from the ChangeEvent class.

Changes apply to references (superClass EReferenceEvent), attributes (...) and resources (...), *You probably need to explain the latter!* as well as object creation. You also need to relate the event class model to EMF metamodel and how tools support model editing (e.g. there are no more change events because these are the only things specific editors allow you to do… or these are the atomic units of all supported model changes -- you note CRUD elsewhere.

In other words: don't just describe: explain rationale and justification, influences, etc.

I would strongly suggest introducing the class structure in one subsection and then starting a new subsection to introduce the operations needed in the prototype. Also, be clear what operations are needed before you start illustrating them.
There's good stuff in here, and some nice explanations e.g. of why EMF constrains what you decide… but presentation needs another iteration or two.

At the end of the section, summarise what you have done, before moving on to "benefits" etc.

§3.5

Start by saying what the section does. It does not make sense just to launch in to "it also does…" which is the gist of the first sentence… so,
This section highlights some of the benefits, and the novel capabilities of the change based persistence prototype…

NB do you need to discuss/summarise how your prototype improves on EMFStore?

§3.6

Without a much more careful reading than I have time for (and note that examiners are also busy people) I cannot work out what the purpose or message of this section is: I can't work out if it is challenges that you met and overcame or challenges that need to be addressed still or things that you will cover later in the thesis… and I am very confused by the "two solutions".
Decide what this section is trying to say. Say what it is trying to say at the start of the section. Make sure that the language clearly relates to either what you have elsewhere in THIS chapter of what you will have later…

If it doesn't make sense to me, it wont make sense to an examiner, even if there is something sensible here and I just missed it by skim reading!

§3.7
Good start… try to talk about what you have in the chapter, not about aims… so how you met an aim etc.

It is probably worth, here and at the start of chapter 3, summarising which RQ(s) and objectives are (partially) met by the chapter

Be clear at the end where you will take the reader in chapter 4….

=========== stopping current review stint: wont get more done for a few days ========