

Annual Report 2016-2017

Lyndon White

February 28, 2018

1 Summary of Research Progress to Date (including any change in focus, and list of publications)

The primary deviation from planning was the request to produce a book for Springer Briefs.

1.1 Publications Arising

1.1.1 Accepted

(Conference Paper) **How well Sentences Capture Meaning** published ADCS 2015. Will be in Part C of thesis.

(Conference Paper) **SOWE2BOW** published CICLing 2016. Will be part of Part B of thesis

(Conference Paper) **SOWE2Sent** published ICDM HDD Workshop 2016. Will be Part of Part B of thesis.

(Conference Paper) **Word-Sense Alignment** CICLing 2018. Will be in Part B of thesis.

1.1.2 Under Review

(Conference Paper) **ColorDist** under-review for ACL 2018. Will be part of Part C of thesis

1.1.3 Pending

(Book): **Neural Representations of Natural Language** In final stage of editing To be published as a Spring Brief. 3 chapters of it will form Part A (lit review) of thesis.

(Conference Paper) **NovelPersective** . First Draft Written. May tweak model slightly again. Some revision to the software required. To be submitted to ACL 2018 Software and Demonstration Track. Will be part of Part C of thesis. Will also be presented to users at the Western Australian Regional Science Fiction Convention (Swancon), as part of there non-archival academic stream.

(Short Journal Paper) **DataDeps.jl** 2nd Draft Written. To be submitted to JMLR Software Track will be part of Part D, or appendix of thesis. This will also be the basic of a presentation “Foundational Tools for Data Driven Research, with Applications in NLP and ML” to be submitted to the (non-archival) Julia Language Conference (JuliaCon).

(Unknown) **Nonnegatively weighted neural networks** I had begun to prepare a theory paper on this, however I discovered it was very similar to an existing work from serveral years early. While there remain some novelties, it may not be worth continuing at this stage as it will not be a nice fit for the thesis topic. Though it does allow for a nice extension to the the ColorDist paper, it would be nearly completely ML focused rather than NLP focused.

1.2 Completion Plan/Deadlines

Key For the symbols in the Timeline

✈ Conference or similar restriction on availability.

★ Externally mandated deadline

☆ Self-imposed deadline

Completion Plan Timeline

- ★9/3/18 Neural Representations of Natural Language Book, submitted to SpringerBriefs.
- ★23/3/18 NovelPerspective Conference Paper submitted to ACL 2018 Software and Demonstration Track.
- 18-24/3/18 CICLing Conference
- 29-2/3/18 Swancon Convention/Easter Break
- ☆15/4/18 DataDeps.jl submitted to JMLR Software Track
This is the last of the works that are currently in high degree of preparedness submitted.
- ☆29/4/18 Dissertation Document Created, with papers all in place, but blank introductions.
- ★30/4/18 JuliaCon Conference abstract due
- ☆14/5/18 Dissertation Part B Introduction/Conclusion written
- ☆28/5/18 Dissertation Part C Introduction/Conclusion written
- ☆17/6/18 Dissertation Part A written based on book
- 15-19/7/18 ACL Conference (if accepted)
 - ☆30/7/18 Dissertation Introduction/Conclusion written
- 7-11/8/18 JuliaCon Conference (if accepted)
 - ★18/8/18 Nomination of Examiners submitted to GRS
 - ★8/9/18 Thesis submitted
- ★17-30/11/18 Potential dates for Oral Defence (If I decide to have one)

1.3 Problems Encountered So Far

- Supervisor Dr Wei Liu left on maternity leave at the end of 2016 until start of 2018. This meant complete loss of her important input on research for 3 months, and only limited ability to get her input for the remainder of the year. This resulted in one project needing to be abandoned after 1 month of work as it required that input; and more generally, Dr Liu's feedback is a very valuable part of the research training program for my studies.
- The NeCTAR allocation which I had been using as the primary compute for my research was due to expire on the 13th of September 2017, and (I was also expectation that I would be able to renew it, as I had previously). However, due to the Pawsey centre's unexpected de-federation, it was instead terminated on the 5th of July 2017, which severely interrupted the middle of a larger series of work. It was not possible to move to the new Nimbus cloud (should I have succeeded in an application to that) as Nimbus does not support the Swift Object Storage which I was making heavy use of. The computations had to be closed off, and I still do not have access to compatible Swift Object Storage. As such, when it becomes necessary to to extend upon that work, time will have to be spent rewriting a nontrivial portion of the research software.
- A more minor but repeated series of interruptions to computation has occurred due to the multiple power failures in the EE building in the last 12 Months. The these have been due to the storm in August, the constructions works from October onwards, and more recently unexplained. Due to failures of communication, even the "planned" power outages were unexpected and resulted in the termination of long running computations.

2 Completion Plan

The gantt chart shown in Figure 1 is updated from that the prior annual report. It shows the tasks remaining to be completed; and the tasks completed.

2.1 Tasks Completed (2016-2017)

As the thesis is by publication, the key progress indicator towards textual completion, in the publications. At this point in total 3 publications have been published, with a 4th currently in review. These publications will form key chapters in the thesis.

For detail on the tasks completed see Section 1.

2.2 Tasks Remaining (2017-2018)

Colour Description (Jan 2017–Jun 2016)

See ??.

Structured Models (Jul 2017–Dec 2017)

This work combined the previously planned Structural Search, and Higher Order models; based on greater understanding. It will consider the possible analysis words in a non-left-to-right structure, for the better capturing of meaning, by considering how each word shifts the possible meanings of the other. This will be considered for a generative application.

The investigation will initially focus on color descriptions, and then will be generalised to apply to arbitrary paraphrases. Both forms of data have ground truth on the similarity of their meanings. In the case of colour descriptions this is in the form of the HSV, HSL, or XY representations. In the case of paraphrases this is only a binary relationship, either a sentence is a paraphrase or it is not.

This work will thus complete the circle of research by linking back to the original work on Semantic Evaluation which focused on determining the similarity of paraphrases.

Dissertation Compilation (Jan 2018 – Mar 2018)

In this time the papers produced during the candidature will be compiled into the thesis. This will involve large scale updating and refining the literature review produced for the research proposal. For this, a lot of information will be extracted from the literature review sections of the papers making up the dissertation, and from the reviews prepared during their investigations.

Dissertation preparation will also involve enhancing the supplementary materials from the prior papers. Due to the tight page limits of the publication venues in this area, most of the papers have associated supplementary material which has been posted online. For example, the paper “Generating Bags of Words from the Sum of their Word Embeddings” has a mathematical proof of the NP-Hardness of the problem, which is as almost long as the paper itself. These materials will either be expanded into chapters, or used to enhance the papers they were supplementing.

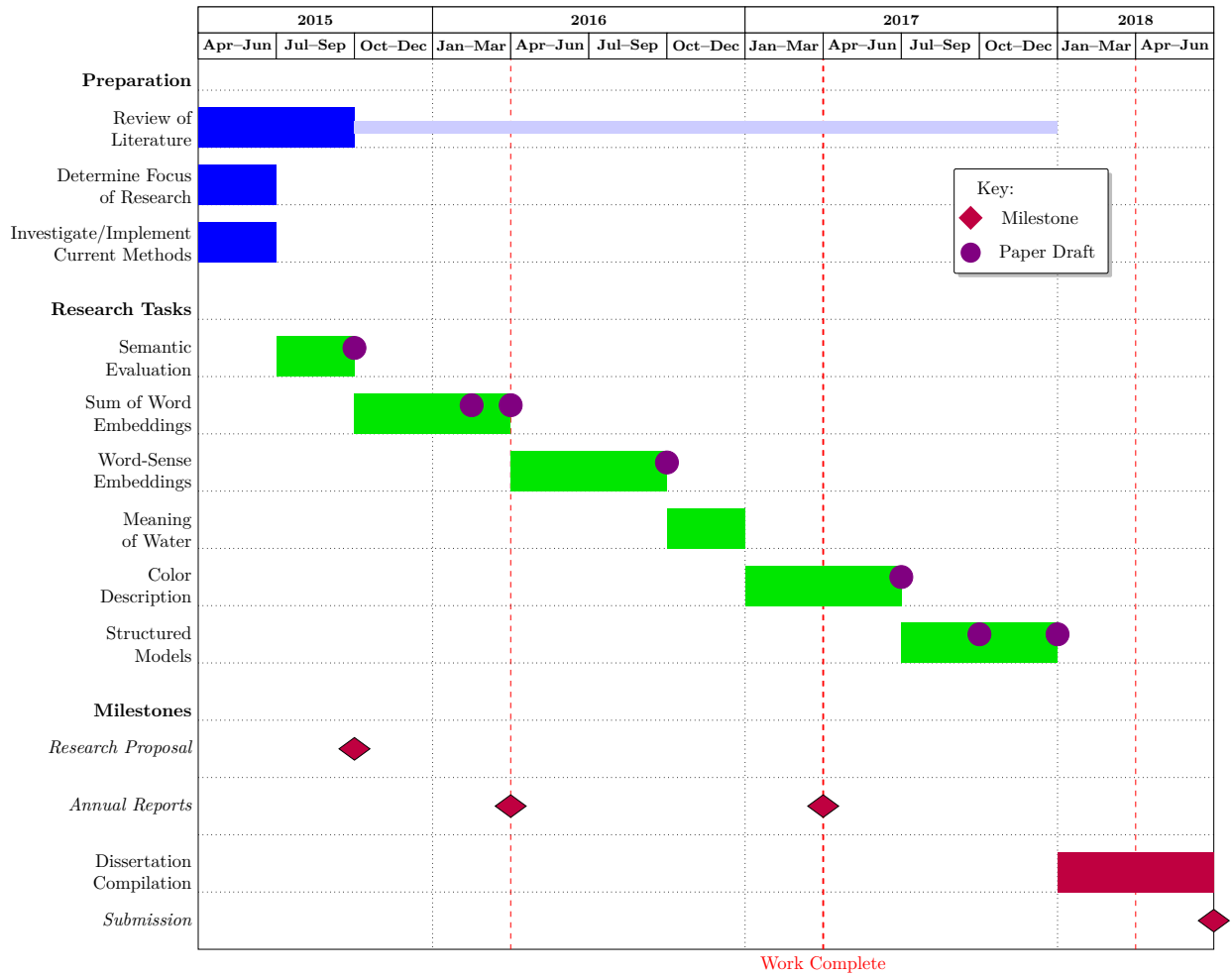


Figure 1: an Updated Gantt Chart