

Bachelor thesis plan

Department of Advanced Computing Sciences
Board of Examiners



Maastricht University

Personal Details

Name: *(Student's family name and first names)*

Ploegstra, Willem

Student number: *(Student's ID number as provided by university)*

6245723

Start date of thesis research

27/01/2023

Expected graduation date: *(Date of the regular bachelor conference where the thesis will be defended)*

30/06/2023

Programme:

BSc Data Science and Artificial Intelligence

☐ The student is enrolled in the KE@Work programme *(check if applicable)*

☐ The student is enrolled in the Marble 2.0 programme *(check if applicable)*

Supervision details

There should be two thesis examiners that approve of this thesis plan. At least one of the examiners acts as a thesis supervisor. The examiners must be DACS staff members with a PhD degree.

Name proposed thesis examiner 1:

Signature:

Steven Kelk

.....

Name proposed thesis examiner 2:

Signature:

.....

The thesis supervisors are DACS staff members that give process based and content-based supervision. At least one of the thesis supervisor acts as an examiner. External supervisors are listed in the next section.

Name DACS thesis supervisor 1: examiner 1

Name DACS thesis supervisor 2: optional

Name DACS thesis supervisor 3: optional

Frequency of contact between DACS thesis supervisor(s) and student:

External collaboration

☐ This thesis involves an external collaboration (*check if an external party is actively involved*)

In case this thesis is not part of an external collaboration proceed to the section “assessment”.

Name external institute: _____ (*Name the institute of external collaboration. In the case of multiple institutes, name the most important one and add an appendix with an overview of external parties*)

This is: another department at Maastricht University

Location external institute:

Name external supervisor:

Frequency of contact between the external supervisor(s)/advisor(s) and student:

Frequency of contact between the external supervisor(s)/advisor(s) and DACS thesis advisor(s):

☐ The student will be hosted by/have a secondment at the external institute (*Check this if the student will work as an internal at the external institute or be visiting the external institute for a period larger than two weeks*)

The student will be hosted by the external institute for:

- ☐ The entire period
☐ A period of _____ weeks

During this period, the work times of the student are as follows:

And the student will spend 40 hours per week on thesis related work.

☐ There is a confidentiality agreement with the following clauses: (*check when applicable*)

☐ The code developed will remain property of the external institute

☐ Special permission has been obtained for limited confidentiality (*See Rules and Regulations and attach documents*)

Note that all thesis must be send to the DACS student affairs office for storing, where they can be labelled “confidential”, but DACS committees and assessment committees (visitatie commissies) will always have (confidential) access for auditing purposes.

The student will receive from the external institute:

- ☐ No financial compensation
☐ An internship fee
☐ Salary
☐ Travel costs reimbursement

On behalf of the external institute, _____ consents with this thesis plan

Signature: _____

Assessment

The assessment will be based on:

(Specify the terms of assessment that at least should contain the thesis, process and the presentation)

- ☒ The contents and form of the thesis, which must at least be sufficient for passing the thesis
 - ☒ A (software) product that accompanies the thesis
 - ☒ The presentation
 - ☒ The process
 - ☐ (optional)
-

Thesis details

Thesis title: *(Working title of the thesis)*

Using Mcts To Generate Upper Bounds For The Tbr Distance

Description of the assignment: *(A single page, describing the assignment and indications of the techniques that will be explored)*

One way to compare two phylogenetic trees is by calculating the TBR distance. Computing the TBR distance is NP-hard to compute. There are already some existing algorithms that are good at calculating exact values and some that can calculate lower bounds fast. However, there are few algorithms that can quickly compute a good upper bound on the TBR-distance. These algorithms are useful since a good upper bound can be used to speed up the computation of exact TBR-distances in addition to giving useful information. [1]

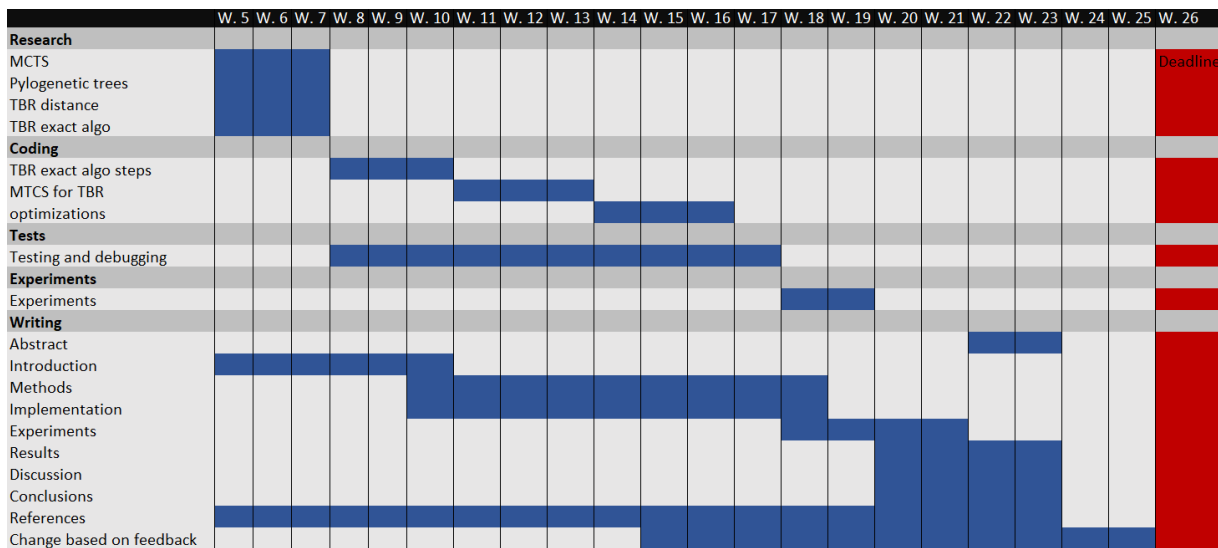
Last year another student from Maastricht University performed work on the same topic and introduced three different algorithms to compute an upper bound on the TBR-distance. The goal of my thesis will be to take these existing algorithms and attempt to improve them to see if the computations of the upper bounds can be improved and/or sped up. The main focus is on one algorithm which used a previously existing exact branching algorithm, which was transformed into a heuristic by stochastically exploring the tree to compute an upper bound. I will attempt to improve the results of this algorithm by exploring the tree multiple times using Monte Carlo Tree Search (MCTS) approach instead of a stochastic one. The expectation is that this will provide better upper bounds.

Additionally, the aim is to optimize the algorithm as much as possible to make it run as fast as possible. This will require an exploration of different data structures that can be used, such as different ways to represent the tree structures needed for the algorithm. Furthermore, some computations/routines that need to be performed in the algorithm will need to be analysed and optimized where possible.

If time permits it I will also attempt to improve one other algorithm created by the student from last year. This algorithm had performance issues and could be improved by making the algorithm more efficient. The algorithm is an iterative greedy approach and thus a increase in performance would make it possible to run the algorithm for more iterations which should improve results while running for the same amount of time.

[1] van Wersch, R., Kelk, S., Linz, S., & Stamoulis, G. (2022). Reflections on kernelizing and computing unrooted agreement forests. *Annals of Operations Research*, 309(1), 425-451.

Planning: *(Provide a complete planning for the thesis period with a reasonable time path for completion, including time for writing the thesis. The preferred format is a Gantt chart that you can attach separately)*



Problem statement: *(single sentence describing the overall challenge faced by the student)*

Create an algorithm that is able to calculate good upper bounds on the TBR distance in a reasonable time.

Research questions: *(provide two to five research questions of a single sentence that by answering them, jointly provide a path for addressing the problem statement)*

1. Does transforming the exact branching algorithm into a MCTS algorithm generate good upper bounds for the TBR distance?
2. How does the MCTS approach compare to the stochastic approach based on time and results?
3. (If time permits) How can the performance of the iterative greedy approach be improved?
4. (If time permits) How does an increase in performance affect the results of the iterative greedy approach?
5. (optional)?

Signature

The student would like to make the following additional statements to the Board of Examiners regarding this thesis plan:

(optional)

Ensure that you: *(Ensure that you check all)*

- ☐ are familiar with the Education and Examination Regulations (EER) that are published on student portal, as is expected of each student
- ☐ are familiar with the Rules and Regulations that are published on student portal, as is expected of each student
- ☐ are familiar with the concept of plagiarism and will refrain from committing it
- ☐ are aware of academic standards, in particular relating to proper citing
- ☐ have filled out the complete form
- ☐ have included a planning and/or attached a Gantt chart with a planning for your thesis
- ☐ have collected the signatures of both examiners and if applicable, the external institute

To avoid students starting too late with their thesis, the deadline for the thesis plans has been set by the thesis coordinator before some of the grades of the preceding semester have come in. Therefore, the possibility of getting conditional approval has been given.

In separate, motivated requests to the Board of Examiners, dispensation from this rule can be requested for cases of hardship. Such motivated requests can be submitted to dke-boe@maastrichtuniversity.nl and are accompanied by a study plan.

☒ I understand that I must have attained 140 ECTS already for this bachelor programme, of which 60 ECTS have been obtained in year 1 and at least 40 ECTS in year 2. I declare that (check one of the following):

☒ I have already obtained the required ECTS and have attached a transcript of my grades showing that

☐ I do not yet have the required number of ECTS, but expect that once the grades of the resits, project/KE@work/MaRBL 2.0 for the semester preceding the thesis semester come in, I will have enough ECTS. Hence, I'm submitting for conditional approval and understand that, if the expected ECTS are not obtained, any conditional approval of the thesis will be automatically withdrawn.

☐ I do not have enough ECTS, but am submitting this plan together with a motivated request for dispensation due to hardship and an individual study plan.

Date:

Student's signature

.....