# CS 4800 Algorithms and Data
# Algorithm Research Project Paper

Professor: Nik Bear Brown
Email: nik@ccs.neu.edu
Office: 451 West Village H
Office Phone: 617 373-5621 (x5621)
Fall 2015

## Topics due:
Research project proposals will be due in week 4 (October 4, 2015) on BlackBoard.

October 4, 2015 – Research topics ideas due.

Upload a paragraph with research project idea, references and option choice to Blackboard by October 4, 2015. This should include some graphics/pictures that illustrate what you want to do.

The topic write-up should be a paragraph or two. You MUST have some links/references as well.

*Note: Only one person needs to upload a project proposal for a group, however ALL group members must upload a text file stating which member uploaded the report and the names of the group members.*

## Progress reports:
Research project progress reports will be due December 1, 2 & 4 2015
Presentations are about 5 minutes.

*Note: Only one person needs to upload a project progress report for a group, however ALL group members must upload a text file stating which member uploaded the report and the names of the group members.*

## Projects Due:
Research project code and reports will be due December 18, 2015.

*Note: Only one person needs to upload a project for a group, however ALL group members must upload a text file stating which member uploaded the report and the names of the group members.*

## Grading Rubric:
The following breakdown will be used for determining the score for the research project:

| Assignment | Points |
| --- | --- |
| Topics/option choice | 100 |
| Progress report | 200 |
| Research Project Report | 500 |

**Research Project:**

Students will be doing empirical algorithm research. Research topics ideas will be due on BlackBoard in week 4, a full in class session presentation of student research projects progress in week 7, research project papers and code due August 22, 2015.

October 4, 2015 – Research topics ideas due.
December 1, 2 & 4 2015 – In class project progress.
December 18, 2015 - Research project papers and code due.

Research project options:

| Option A | Percent of Total Grade |
| --- | --- |
| Group Project | 10% |

| Option B | Percent of Total Grade |
| --- | --- |
| Individual Project | 10% |

Note:  Only one person needs to upload a project for a group, however _ALL group members must upload a text file stating which member uploaded the report and the names of the group members._

## Research Projects
CS 4800, Algorithms and Data, will have individual and/or group research papers.  These papers should be in a style that could be submitted to a conference, workshop or journal. Students should begin to work on this in week 1.

These assignments will provide practice in real-world analysis and application of algorithms. One project is a group of one to three people and one by yourself. In each project, you or your group will write a short (4-8) page paper related to algorithms. The research can take one of the following forms:

    i.    Tweaking an existing algorithm.
    ii.    Applying an existing algorithm in a novel context.
    iii.    Validating an existing algorithm in real-world contexts.
    iv.    Creating a novel algorithm.

### _Topic 1_
**Tweaking an existing algorithm**
This project involves finding an existing algorithm and making a modification that might improve its recall, precision, scalability, memory usage or speed. Other tweaks may include changing the algorithm so it applies in a different context.  For example, tweaking the Gale-Shapley stable marriage algorithm to include same sex couples.

### _Topic 2_
**Applying an existing algorithm in a novel context**
This project involves finding an existing algorithm and applying it in a novel way. This might involve applying genetic sequencing algorithms to the optimization of hardware resource sharing, or the use of genetic algorithms (GAs) to optimize the arrangement of class schedules.  This topic is very closely related to "Tweaking an existing algorithm" in that one usually needs to make changes to use an algorithm in a novel context.

### _Topic 3_
**Validating an existing algorithm in real-world contexts**
This project involves finding how practically useful and applicable an existing algorithm really is in the "real world." This involves empirically validating the sensitivity, recall, precision, scalability, memory usage and speed claims of an algorithm with realistic, noisy and non-ideal data.

**Creating a novel algorithm**
This project involves creating a novel algorithm that can answer an interesting real-world question. This topic is very closely related to "Tweaking an existing algorithm" in that one usually extends and improves what exists rather than create totally from scratch.

**Topic versus Thesis**
A topic is a general interest. A thesis statement presents an assertion; what you intend to do and how you intend to prove/convince others what you did is correct.  A topic of interest might be "What keywords should I add to my tweet to make it more viral?"

A thesis is more specific and is framed such that the assertion is testable.  A thesis would be "We believe adding keywords according to our algorithm will significantly improve a tweets virality."  The paper would then involve quantitatively defining virality and comparing the random tweets and the tweaked tweets with the keyword algorithm using reasonable real-world data (e.g. Twitter).

**What you need to do**
 A. Find a group
 B. Find topics
 C. Upload a paragraph with individual and group topics to Blackboard by Week 4 (June 4, 2015).
 D. Do research, find data, and write code.
 E. Research project progress reports will be due in Week 6 (June 17 & 18 2015)
 F. Write quality papers.
 G. Research project code and reports will be due June 29, 2015.

The research project papers must have:
 a. Abstract (5 %)
 b. Introduction (5 %)
 c. Code with Documentation (50%)
 d. Results (25 %)
 e. Discussion (10 %)
 f. References (5 %)

**Algorithms "State of the Art" Research Paper**
This assignment will provide practice in computer Algorithms research. You will write a short (4-8) page paper on the "State of the Art" some topic in computer Algorithms.  The paper should take the form of a "literature review" paper:

The state of the art paper must have:
 a. Abstract (5 percent)
 b. Introduction (5 percent)
 c. State of the Art Review (80 percent)
 d. Discussion (5 percent)
 e. References (5 percent)

**What is a "State of the Art" Research Paper?**

A "state of the art" research paper is a short instructional paper related to some interesting aspect of computer Algorithms. A "state of the art" review of literature papers are presented to describe and explain cutting edge techniques or procedures used in computer algorithms. The technique should be relatively new or not widely understood or used in practice or super interesting. The goal of the paper is to give Algorithms artists and scientists the information they need to critically evaluate the pros and cons of a technique and the details to implement it in their practice if they choose.

*What is a "state of the art" review of literature?*

The review should be a self-contained unit. While references are critical some explanation of the background needs to be in the paper. The type of review is a required as part of grant and research proposals and often a chapter in theses and dissertations. Generally, the purpose of a review is to analyze critically a segment of a published body of knowledge through summary, classification, and comparison of prior research studies, reviews of literature, and theoretical articles.

*The introduction*

In the introduction, you should:
 i. Define or identify the general topic, issue, or area of concern, thus providing an appropriate context for reviewing the literature.
 ii. Point out overall trends in what has been published about the topic; or conflicts in theory, methodology, evidence, and conclusions; or gaps in research and scholarship; or a single problem or new perspective of immediate interest.
 iii. Establish the writer's reason (point of view) for reviewing the literature; in essence answer the question of why this subject is interesting.

*Writing the body, the "state of the art" review*

In the body, you should:
 i. Group research studies and other types of literature (reviews, theoretical articles, case studies, etc.) according to common themes of methods.
 ii. Summarize individual studies or articles with as much or as little detail as each merits according to its comparative importance in the literature, remembering that space (length) denotes significance.
 iii. Provide the reader with strong "umbrella" sentences at beginnings of paragraphs, "signposts" throughout, and brief "so what" summary sentences at intermediate points in the review to aid in understanding comparisons and analyses.

*Writing the discussion*

In the discussion, you should:
 i. Summarize major contributions of significant studies and articles to the body of knowledge under review, maintaining the focus established in the introduction.
 ii. Evaluate the current "state of the art" for the body of knowledge reviewed, pointing out major methodological flaws or gaps in research, inconsistencies in theory and findings, and areas or issues pertinent to future study.
 iii. Conclude by providing some insight into guiding the reader with recommendations.

*Writing the references*

In the references, you should:
  i.    Be exhaustive
  ii.   Any source information that you provide in-text must correspond to the source information on the Works Cited page.
  iii.  Use a consistent citation style (e.g. MLA, APA, Chicago, and CSE citation styles http://www.apastyle.org/ , http://www2.liu.edu/cwis/cwp/library/workshop/citation.htm, http://www.lib.berkeley.edu/how-to-find/cite-sources ).


**Final Report Format**

500 points possible

| Requirement | Points |
|---|---|
| Abstract | 5% |
| Introduction | 5% |
| Methods/Instructions/Code | 50% |
| Results | 25% |
| Discussion | 10% |
| References | 5% |


**Submission**
You will submit your assignments via BlackBoard.
Click the title of assignment (blackboard -> assignment -> <Title of Assignment>), to go to the submission page.

**What Topics are Interesting?**
- Sequence alignment
- Secure Communication
- Scheduling (courses, trains, etc.)
- Packing (where to best place items in store/website/warehouse)
- Path finding (who/what is closet? How do I get there? Extracting a Retweet's Origins?)
- What terms are associated with your Twitter/ Facebook/ LinkedIn/ Google+:/ GitHub/ Web handle?
- Which genes cause cancer?
- What users are associated with your Twitter/ Facebook/ LinkedIn/ Google+:/ GitHub/ Web handle?
- Friendship cliques on Twitter/ Facebook/ LinkedIn/ Google+:/ GitHub/ Web
- Sentiment analysis/influence analysis. Who are most influential on Twitter/ Facebook/ LinkedIn/ Google+:/ GitHub/ Web?
- Mapping (jobs, housing, crime, etc.)
- At what price should I start an Ebay auction?
- Do I want to follow that person back?
- Where should I place transmission towers?
- What keywords should I add to my tweet/post?

- Identification of Transcription Factor Binding Sites
- When should I tweet/post?
- Matching (Who is like/unlike me? What is the best TV, college, etc. for me?)
- What is the reach of my tweet/post?
- What are people saying about me on Twitter/ Facebook/ LinkedIn/ Google+:/ GitHub/ Web?
- Should I add a picture/url to my tweet/post?

There are an infinite number of interesting topics to which algorithms could be applied.  You want to keep your topic simple and doable. A topic like "Which genes cause cancer?" is of great interest but too broad for a topic. Analyzing existing gene expression quantification algorithms is more suitable for a project. The purpose of these projects is for students do get their "hands dirty," not to necessarily develop break-through algorithms.  We will discuss how to do this research at the end of week 1 and beginning of week 2 in class.