SELECTION AND APPROVAL OF PROJECTS

MEETING TO DISCUSS YOUR PROPOSED PROJECT DURING FIRST HALF OF THE COURSE

You are to select a project in an area that interests you. Please meet with me (during scheduled office hours or at another mutually agreed time) prior to **Wednesday, October 29, 2003** to discuss your proposed project (or your list of proposed alternative projects). The meeting will be at the Medical School Office Building (MSOB) (corner of Welsh Road and Campus Drive). I will circulate a sign-up sheet in class for appointments. If you are watching on TV or web from a distance, send me an e-mail message at Koza@Stanford.Edu to arrange an appointment on campus. If you are not in the Silicon Valley area, send me an e-mail and we'll have to work out your project by e-mail and phone. These meetings typically occur in the third or fourth week of the class, but they can occur sooner. The lectures in the first half of the class are designed to survey many different areas of genetic algorithms and genetic programming so as to help you think of a project topic. If you have never been in a course where a self-selected project is a major part of the course or are an undergraduate, we should meet with you no later than the third week of the course (as opposed to the last mine). In most cases, we will agree on a project during our first meeting. Usually, we will have at least one meeting in the second half of the course to discuss progress on your project.

PROJECT APPROVAL

I will either approve your proposal, suggest modifications or alternatives, or arrange a further meeting to discuss the project with you. Please turn in the PROJECT DESCRIPTION sheet (below) on paper or by e-mail to me at koza@cs.stanford.edu after we reach agreement.

SOFTWARE RESOURCES

Your project may use any available GA, GP, or other applicable software, including, but definitely not limited to,

- the GENESIS software in C for GA,
- Dave's DGPC software (1-byte version) for GP,
- Goldberg's Simple Genetic Algorithm (SGA) in PASCAL in Goldberg's book,
- the LITTLE LISP GP software in Common LISP as described in Appendix B of *Genetic Programming* book, or

There are upwards of a dozen other versions of GA and GP software available on the web.

Projects typically involve using existing software, modifying it, and writing additional software to accompany the existing software. It is definitely not recommended that you attempt to re-create software to implement basic GA or GP.

HARDWARE RESOURCES

Your project may use any computer hardware available to you. This will often be your own computer, a computer available on campus at the various clusters, or a computer available to you as part of employment.

INTERMEDIATE MEETINGS DURING SECOND HALF OF THE COURSE

After approval, I like to meet with you (during scheduled office hours or at another mutually agreed time) during the course of your project to review progress, discuss roadblocks and milestones, and redefine the project based on preliminary good or bad results. There is no requirement for such additional meetings; however, I urge you to do so at least once during the course of doing your project (i.e., during the second half of the course).

WRITTEN REPORT ON THE PROJECT

Seven copies of the final, camera-read copy of your written 10-page paper describing your project are due **Wednesday June 5, 2002 at 11 AM (No Extensions!!!).** Note that four copies will be sorted into envelopes of other students during this final lecture, so **it is imperative that the papers be turned in at the beginning of this lecture — at 11 AM — not in the middle or end**. I'll make special individual arrangements, individually, with TV/web students who do not come to campus for this last lecture.

BMI 226 / CS 426
PROJECT DESCRIPTION SHEET
DUE Wednesday, October 29, 2003

NAME_____

Date: _____

(1) Statement of the problem (in the form of the anticipated likely abstract of the final paper):

(2) Type of algorithm:

___ Fixed-length genetic algorithm (GA)
___ Variable-length genetic algorithm (VGA)
___ Genetic Algorithm operating on specialized data structure
___ Genetic programming (GP)
___ Developmental GP (cellular encoding)
___ Messy GA (mGA)
___ Learning classifier system (LCS)
___ machine assembly language GP
___ Other:_____

(3) Computer that you intend to use for experiments (if any):

(4) Software you intend to use:

___ GENESIS in C for GA
___ LIL GP for GP
___ ECJ for GP
___ Dave's DGPC in C for GP
___ Goldberg's SGA in Pascal for GA
___ Goldberg's SCS in Pascal for CS
___ LITTLE LISP GP code for GP
___ Other:_____

(5) Word processor you plan to use for your paper

___ Microsoft WORD
___ Latex
___ Other _____

BMI 226 / CS 426

FORMAT OF PAPERS FOR PROJECTS

DEADLINE FOR PAPERS: Wednesday December 3, 2003 at 3:15 PM (No Extensions!!!)

NUMBER OF COPIES: Please bring 7 stapled copies of your paper to class (and 11 in the very unlikely event that you are doing a 2-person project). Note that four copies will be sorted into envelopes of other students during this final lecture, so **it is imperative that the papers be turned in at the beginning of this lecture — not in the middle or end**. TV/web students who do not come to campus for this last lecture should send in paper in the usual way so that it arrives by the morning of the last class. In special cases, we can arrange so that it can be sent by fed-ex to arrive by 10:30 AM at my house.

LENGTH OF PAPER: The paper is not to exceed 10 pages. The 10-page limit includes EVERYTHING. That is, this page limit includes all of the usual elements of a technical paper for a journal or conference, namely

- the title,
- author's name and address (i.e., full physical address, e-mail address, and phone),
- the abstract (75-200 words),
- the full text of the paper, including all section and subsection headings, subheadings, figures (if any), and tables (if any),
- acknowledgments (if any),
- the bibliography, and
- the appendix of the paper (if any).

Your paper should stand alone and should not refer to your computer program listing in any way. In the unlikely event that it is necessary to show specific parts of your computer program in your paper, then that part of the listing must be entirely contained within the 10-page limit of your paper (e.g., as a table in the paper, in the running text, or the appendix).

FORMAT OF PAPER: The paper is to be CAMERA-READY. The paper must be SINGLE-SIDED. Please use a good quality laser printer, since the original will be used for printing. The format is

- single-spaced,
- 10-point type,
- 8 1/2" x 11" paper,
- one or two columns per page,
- FULL 1" margins at top, bottom, left, and right. Only the page number can be outside this margin. • pages numbered from 1 to 10 at the TOP LEFT of each page outside this margin, and
- the full text of the paper continues on the same page right after the title, author's name and address, and abstract (i.e., there is no separate title page or cover page).

Your paper should employ a style consistent with the above requirements and resemble the sample paper below. In citing items in your paper, please use the "(Smith 1990)" author-date style, rather than the "consecutive number" style. Your paper should look like the abbreviated sample paper below.

The only thing that will be done to your paper in preparing it for the book is to white-out your page number (from 1 to 10) and add the consecutive page number so that all the papers end up consecutively numbered within the book.

PROCEDURE: The single-sided camera-ready copy will be used to print your paper. One copy goes to the instructor for grading. One other copy will go to the TA for grading. The other 4 copies will be sorted during the last meeting of the class into random batches containing 4 different papers each (not your own, of course). The batches will be distributed at the end of the last meeting of the class to each student submitting a paper at the beginning of the hour. In addition, each batch will contain 4 blank copies of the individual paper review form and 1 blank copy of the overall review form. The 4 copies of the review form and the 1 overall review form are the take-home final for this course. The exam is individual and subject to the honors code.

THE PEER-REVIEW PROCESS: Start by reading the 4 papers in your batch. The 4 individual paper review forms ask you to make comments on the strengths and weaknesses of each paper (viewed alone) in much the same manner as researchers from the academic, commercial, government, and R & D worlds are asked to evaluate papers submitted for publication in technical journals and books and for presentation at conferences. In addition, the 1 overall review form asks you to make an overall comparative ranking of the 4 papers and give specific comparative reasons for your ranking (i.e., A is better than B because ... and B is better than C because...). Please use the last 3 digits of your student ID, driver's license, or social security number as your reviewer number.

Keep in mind that your reviews will be graded in terms of the quality of the SPECIFIC REASONS you give for your specific comments on the individual papers and the quality of the SPECIFIC REASONS you give for your overall ranking of the papers. Thus, it does not matter whether your random batch happens to contain unusually good or bad papers. Unexplained comments, such as "good paper" are of no interest. We are interested in specific comments supported by specific reasons.

DEADLINE FOR TAKE-HOME FINAL: Your 4 reviews and your 1 overall review form are due at the time shown above.. There is no need to return the 4 papers you are reviewing. However, remember that these papers and any ideas in them belong to the student who prepared the paper.

RETURN OF PAPERS: The 4 review forms containing other students' comments about your paper will be returned to you in an envelope sometimes after the deadline for submitting grades to the University. An e-mail message will be sent to the mailing list for the class announcing the time when they are placed there.

BINDING OF PAPERS INTO BOOK FORM: Unless you specify otherwise (by writing “NOT FOR INCLUSION INTO BOOK” on the top line of the first page of your paper), your paper will be bound into a book and distributed to all the students in the class who turn in a paper. The book will NOT be copyrighted. The book will specifically note that each student retains all rights to his or her paper. By submitting a paper for the book, you agree to its publication in the book and the making of whatever additional extra copies of the book may be made for sale by the Stanford University Bookstore or distribution by the instructor.

IDEAS FOR CONTENTS OF THE PAPER

The written report on the project should address each the following issues and questions (unless clearly inapplicable).

1. STATEMENT AND MOTIVATION OF THE PROBLEM: In the introduction or elsewhere, there should be a crisp statement of the problem being handled by the paper. In addition, you should include some kind of (less formal) motivation as to why anyone would be interested in the problem in the first place. Usually this motivation goes into the introductory section. This should position your problem in the grand scheme of the things.

2. MAJOR PREPARATORY STEPS: Some part of the paper should detail the methods you used in doing the work (perhaps in a section entitled "methods"). Your paper should identify the four major preparatory steps for a genetic algorithm application or the five major preparatory steps for a genetic programming application. Include a tableau as a table itemizing these choices for at least one version of your problem (e.g., the first one discussed in the paper).

If the project involves genetic classifier systems, describe (i) the detectors (representation scheme), (ii) effectors (including conflict resolution), (iii) message list (contents, size) (iv) condition part(s) of the classifiers, (v) action part of the classifiers, (vi) the method for allocation of credit, (vii) the method of awarding environmental reward, and (viii) all the parameters being used.

3. SPECIFIC EXAMPLES OF THE STRUCTURES UNDERGOING ADAPTATION: To make the previous step concrete, give two specific illustrative examples of possible individual structures in the population. Give a verbal interpretation of that structure in the terminology of the problem domain. Say what the fitness of the two example individuals is. This is very helpful for the reader in understanding exactly what you are doing.

4. SIZE OF THE SEARCH SPACE: This step may or may not be possible if you are using GP. It is usually possible if you are using GA. If possible, give a count (or approximation or estimate) of the total number of possible structures in the search space involved. Depending on the difficulty of the combinatorics of the problem you selected, you may or may not be able an exact mathematical expression counting these structures. If you cannot give an exact count, give an upper bound, lower bond, a reasonable estimate of the count, or, if none of the foregoing are possible, give a count for a special case.

5. THE FITNESS MEASURE:

(a) How is the range of the fitness measure guaranteed to end up in the range [0,1] with the better values being smaller and the best value being zero? (This may be so obvious that is not worth mentioning, but it should be included if it is not dead obvious).

(b) Identify any constraints or penalties involved for infeasible structures.

6. THE ENVIRONMENT (FITNESS CASES):

(a) What is the environment (fitness cases)?

(b) Why is this environment is sufficiently representative of the problem as a whole so that computing fitness using this environment will likely lead to a solution applicable and generalizable to the problem as a whole?

7. THE GENETIC OPERATORS:

(a) What operators act on the structures to modify them?

(b) If your operators are not the ordinary operators of reproduction, crossover, and mutation, (i) state the number of structures on which it acts, (ii) give a specific illustrative example of each unusual operator, (iii) state why the operator is well-defined and guaranteed to produce another allowable structure for any choice of structure(s) from your set of possible structures.

(c) If your operators do not always produce allowable structure(s) for some structure(s) on which it may act, what is then done (e.g. penalties for constraint violation, repairing, discarding)?

8. ACCESSIBILITY OF THE ENTIRE SEARCH SPACE:

(a) Do the operators allow all structures in the space of all possible structures to be created (either initially or later)?

(b) Give a reason why this is true (or not true).

(c) If this is not true, what limitations do the operators have and give a reason why this limitation on the operator's ability to create all possible structures is reasonable (and what effects the limitation has).

9. THE ADAPTIVE PLAN:

(c) Why did you choose each of the most important parameters that controls the adaptive plan?

10. THE SCHEMATA:

(a) What are the schemata that are involved in the processing involved? If this is not clear, discuss what the schemata might be and try to say what they are for a limited special case of some kind.

(b) Give a count of the total number of schemata. If that is too hard for your particular problem (e.g., GP), give an upper or lower bound or other reasonable estimate on the number of schemata involved for the general case or a special case.

(c) What is the minimum and maximum number of schemata that are processed at each time step (by each operator)?

11. THE POPULATION:

(a) What is the size of the initial population of structures?

(b) How did you arrive at your choice for this number?

(c) How is the initial population of structures created?

(d) Is any domain-specific heuristic knowledge used to "prime" the initial population (with either whole structures or parts)?

12. REPLICATION OF RESULTS:

Identify all the other equipment, software, hardware, etc. associated with your experiment so that you provide enough information so that someone could replicate your experiment in his own "laboratory" and (presumably) get the same results (within the limits inherent in any probabilistic process). Give some idea of how much computer time is involved for some specifically cited run.

13. RESULTS AND OBSERVATIONS:

(a) What did you observe in your experiment(s)?

(b) What happened in the beginning, intermediate, and final stages of at least one run and what is the interpretation?

14. DISCUSSION AND INTERPRETATION:

What is the interpretation for what you observed?

15. PROBLEMS ENCOUNTERED AND FAILURES:

(a) Did your experiment run into any problems associated with loss of population diversity or premature convergence at any stage? How could you solve that problem?

(b)Did your experiment run into any problems associated with getting trapped on local optima? How could you solve that problem?

(c) Did your experiment contain any problems associated with deception? How could you solve that problem?

(d) What false starts did you try (and find did not work). (This is a much neglected, but very important, part of reports on research).

16. CONCLUSION:

What overall conclusions can be reached from what you did and observed? This section should summarize what the paper actually demonstrated. It is not a speculative (future works) section (see below).

17. FUTURE WORK:

What further study might be the logical follow-up for whatever you did if someone wanted to further pursue your study?

Simultaneous Discovery of Detectors using Genetic Programming

John R. Koza
Stanford, California
koza@stanford.edu
http://www.smi.stanford.edu/people/koza/

Abstract: This paper describes a technique for simultaneously discovering detectors and a way of combining the detectors to solve a problem of pattern recognition using genetic programming. The technique uses automatically defined functions. Automatically defined functions enable genetic programming to define potentially create useful functions dynamically during a run and to facilitate a solution of a problem by automatically and dynamically decomposing the problem into simpler subproblems. The technique is illustrated with a problem of letter recognition.

1. Introduction and Overview

Conventional approaches to problems of pattern recognition and machine learning usually require that the user hand-craft detectors for key features in the problem environment. Conventional approaches often additionally require the user to specify in advance the size and shape of the eventual way of combining the detectors into a compete solution.

It is desirable, however, ...

Section 2 of this paper provides background and reviews previous work in this field. Section 3 of this paper states the problem. Section 4 describes the methods. Section 5 states the results. Section 6 discusses the results. Section 7 states the conclusion. Section 8 discusses future work.

2. Background

3. Statement of the Problem

4. Methods

The "methods" section provides the details of how you went about doing whatever you did.

Papers in this course **must** contain a **tableau** summarizing the choices made for at least one version of the problem being discussed. A typical GA tableau is shown below. Table 1 summarizes the preparatory steps for the 10-member truss problem. If GP is being used, the "representation scheme" would, of course, be replaced by sections on the terminal set and function set. If the GP programs have multiple branches, there would separate sections on the terminal set and function set for each distinct type of branch.

Table 1. Tableau for the 10-member truss problem.

Objective:	
Representation scheme:	• Structure = • K = • L = • Mapping from points in search space of the problem to structures in the population =
Fitness cases:	
Fitness:	
Parameters:	Population size M =
Termination criteria:	
Result designation:	

5. Results

The "results" section states your results.

Most papers have one or more figures showing something about the set-up of the problem or the results. **"Tables" differ from "figures" in that tables consist entirely of text while figures have graphical elements.**

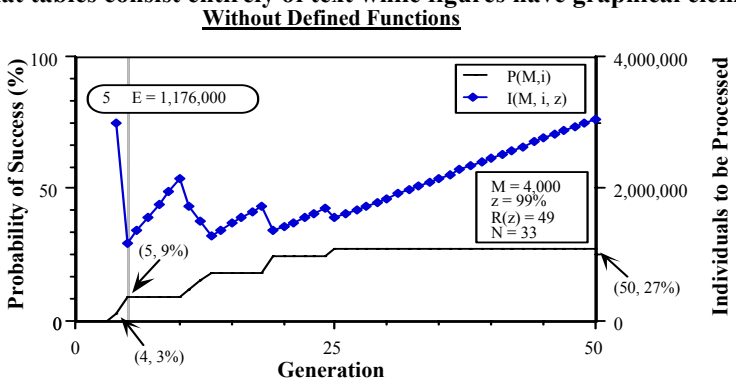

Figure 1 Performance curves for the two-boxes problem.

Table 1 Fitness of best-of-run individual for the two-lag plant problem.

	Step size (volts)	Internal Gain, K	Time constant, τ	Fitness
0	1	1	1.0	0.0220
1	1	1	0.5	0.0205
2	1	2	1.0	0.0201
3	1	2	0.5	0.0206
4	10^{-6}	1	1.0	0.0196
5	10^{-6}	1	0.5	0.0204
6	10^{-6}	2	1.0	0.0210
7	10^{-6}	2	0.5	0.0206

6. Discussion of Results

It is desirable to separate the "results" from the "discussion" and interpretation of the results. In particular, by having a separate "discussion" section, you will be reminded to do some "discussing."

7. Conclusion

This section of the paper summarizes what the paper actually did. It might being, " This paper has demonstrated …" The conclusion section is usually short. It often somewhat resembles the contents of the abstract.

8. Future Work

The "future work" section is not to be confused with the "conclusions: of the paper. The "conclusion" section states what is **actually done** in the paper. The "future work" section speculates on possibly interesting future work that might be done by the author or the reader at some time in the future.

9. Acknowledgements (Include only if there are any)

References

STYLE NOTE: The 3 references below illustrate a good style for a book, a paper in a collection of papers (e.g., conference proceedings or an edited book consisting of chapters contributed by different authors), and a journal article. You should use a consistent style for each type of item in a bibliography. I

Holland, John H. 1975. *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press.

Koza, John R. 1992. Evolution of subsumption using genetic programming. In Varela, Francisco J., and Bourgine, Paul (editors). *Toward a Practice of Autonomous Systems: Proceedings of the first European Conference on Artificial Life*. Cambridge, MA: The MIT Press. Pages 110-119.

Newman, R. C. 1988. Self-reproducing automata and the origin of life. *Perspectives on Science and Christian Faith*. 40(1) 24-31.